(12) UK Patent Application (19) GB (11) 2 307 477 (13) A

(43) Date of A Publication 28.05.1997

(21) Application No 9624453.8

(22) Date of Filing 25.11.1996

(30) Priority Data
(31) 9523959 (32) 23.11.1995 (33) GB
9525555 14.12.1995
9617961 28.08.1996

(71) Applicant(s)
Cancer Research Campaign Technology Limited
(Incorporated in the United Kingdom)
Cambridge House, 6-10 Cambridge Terrace, Regent's Park, LONDON, NW1 4JL, United Kingdom
Duke University
(Incorporated in USA - North Carolina)
230 North Building, Research Drive, Durham, North Carolina 27708, United States of America

(51) INT CL[6]
C12N 15/11 , A01K 67/027 , A61K 31/505 38/17 39/395 48/00 , C07K 14/82 , C12N 15/12 15/89 , C12Q 1/68 , G01N 33/574

(52) UK CL (Edition O)
C3H HB7P HB7T H645 H656 H674
C6Y Y419 Y501 Y503
U1S S1287 S1303 S1313 S1332 S1334 S1337 S2419

(56) Documents Cited
Science 1994,265,2088-2090 Nature Genetics 1996,12,333-337 Oncogene 1996,13,213-221 Proc.Natl.Acad.Sci.USA 1996,93,690-694

(58) Field of Search
UK CL (Edition O) C3H HB7M HB7P
INT CL[6] C07K 14/82 , C12Q 1/68 , G01N 33/574
ONLINE: WPI,CLAIMS,DIALOG/BIOTECH

(72) Inventor(s)
Phillip Andrew Futreal
Richard Francis Wooster

(72) and (74) continued overleaf

(54) **Materials and methods relating to the identification and sequencing of the BRCA2 cancer susceptibility gene and uses thereof**

(57) The identification and sequencing of the BRCA2 gene is disclosed as well as the amino acid sequence of the corresponding BRCA2 polypeptides. BRCA2 alleles including those with mutations in the BRCA2 gene which are associated with a predisposition to develop cancer, especially breast and ovarian cancer are also disclosed. The present invention further relates to polypeptides encoded by the above nucleic acid. The present invention further relates to uses of such BRCA2 nucleic acid and BRCA2 polypeptides in particular in the diagnostic, prognostic or therapeutic treatment of cancer.

GB 2 307 477 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

(72) cont
Alan Ashworth
Michael Rudolf Stratton

(74) Agent and/or Address for Service
Mewburn Ellis
York House, 23 Kingsway, LONDON, WC2B 6HP,
United Kingdom

Fig.1.

Sequence 1. The BRCA2 cDNA

The sequence is shown 5' to 3'

NCNAAAGTTTAAGGGAGTTGTTTAGGAGGAAATTTGATTTAATTCAGAAA
NTGAGCATAGTNTTCCANTATTCACCTACGTTTTAGACCAAAATGTATCA
AAAATACTTCCTTCGTGNTGATAAGAGAAACCCAGAGCACTGTGTAAACT
CCAGAAATGGAAAAAACCTGCAGTAAAGAATTTAAATTATCAAATAACTT
AAATGTTGAAGGTGGTTNTTCAGAAAATAATCACTCTATTAAAGTTTCTC
CATATCTCTCTCAATTTCAACAAGACAAACAACAGTTGGTATTAGGAACC
AAAGTNTCACTTGTTGAGAACATTCATGTTTTGGGAAAAGAACAGGCTTC
ACCTAAAAACGTAAAAATGGAAATTGGTAAAACTGAAACTTTTTCTGATG
TTCCTGTGNAAACAAATATAGAAGTTTGTTCTACTTACTCCAAAGATTCA
GAAAACTACTTTGAAACAGAAGCAGTAGAAATTGCTAAAGCTTTTATGGA
AGATGATGAACTGACAGATTNTAAACTGCCAAGTCATGCCACACATTNTC
TTTTTACATGTCCCGAAAATGAGGAAATGGTTCTGTCAAATTCAAGAATT
GGAAAAAGAAGAGGAGAGCCCCTTATCTTAGTGGGAGAACCCTCAATCAA
AAGAAACTTATTAAATGAATTTGACAGGATAATAGAAAATCAAGAAAAAT
CCTTAAAGGCTTCAAAAAGCACTCCAGATGGCACAATAAAAGATTGAAGA
TTGTTTGTGCATCATGTTTCTTTAGAGCCGATTACCTGTGTACCCTTTCG
CACAACTAAGGAACGTCAAGAGATACAGAATCCAAATTTTACCGCACCTG
GTCAAGAATTTCTGTCTAAATCTCATTTGTATGAACATCTGACTTTGGAA
AAATCTTCAAGCAATTTAGCAGTTTCAGGACATCCATTTTATCAAGTTTC
TGCTACAAGAAATGAAAAAATGAGACACTTGATTACTACAGGCAGACCGA
CCAAAGTCTTTGTTCCACCTTTTAAAACTAAATCACATTTTCACAGAGTT
GAACAGTGTGTTAGGAATATTAACTTGGAGGAAAACAGACAAAAGCAAAA
CATTGATGGACATGGCTCTGATGATAGTAAAAATAAGATTAATGACAATG
AGATTCATCAGTTTAACAAAAACAACTCCAATCAAGCAGCAGCTGTAACT
TTCACAAAGTGTGAAGAAGAACCTTTAGATTTAATTACAAGTCTTCAGAA
TGCCAGAGATATACAGGATATGCGAATTAAGAAGAAACAAAGGCAACGCG
TCTTTCCACAGCCAGGCAGTCTGTATCTTGCAAAAACATCCACTCTGCCT
CGAATCTCTCTGAAAGCAGCAGTAGGAGGCCAAGTTCCCTCTGCGTGTTC
TCATAAACAGCTGTATACGTATGGCGTTTCTAAACATTGCATAAAAATTA
ACAGCAAAAATGCAGAGTCTTTTCAGTTTCACACTGAAGATTATTTTGGT
AAGGAAAGTTTATGGACTGGAAAAGGAATACAGTTGGCTGATGGTGGATG
GCTCATACCCTCCAATGATGGAAAGGCTGGAAAAGAAGAATTTTATAGGG
CTCTGTGTGACGTAAAGGCCACATAGTGGATAAACCATGTGCCACTTGAG
GAATTGAGAAAAAGTCAGTGTGTTTAGAACATAAGGAGAGGGGCAGAGAA
ATGGATCTGGTACAGGAGAATCAATGAGACCGGGTAAGAAACAGAAAAGG
GGCTGCACCAAATGATTGACACGCTCTGCAAACTTCTTTTGTTGGCTCAA
GTTGTGGCTCAAGAGGTGAGAAGGTAAGGCCAGACTTATTTGTTATGATT
TGCCCTTTAAATTGAAGCCTTAAGATTGGAATTCGATATCAAGCTTATCG
ATACCGTCGACCTCGAG

Fig.2(a).

Genomic organisation of the BRCA2 gene

exon29f

genomic cgtaaaaatggaaattggtaaaactgaaacttttctgatgttcctgtga

genomic anncaaanatagaagtttgttctacttacncccaagattcagaaaactac

genomic tttgnaacagaancagnagaaattgctaaagcttttatggaagatgatga

genomic actgacagattctaanctcccaantcatgccacacattctctttttacat

cDNA GTCCCGAAAATGAGGAAATGGTTTAGTCAAATTCAAGAATTGGAAAAAGA
genomic GTCCCGAAAATGAGGAAATGGTTTTGTCNAANTCAAGAATTGGAAAAAGA

cDNA AGAGGAGAGCCCCTTATCTTA-GTGG
genomic AGAGGAGAGCCCCTTATCTTANCNGGGtaagtnttcatttttacctttcgcc

genomic nttgccaatcactatttttaaagtgtttattcagtagacttggtatgcta

genomic acaattaagagtgttataaactatgtcttttcagccatttttgtgtagtc

genomic agtttgggggagtatggtttgatatacagatacacagattcagtattcgt

genomic atacagatttgatatcttggtatacagattcgatatctctgaatctgtat

Fig.2 (b).

exon29e

cDNA GAGAACCCTCAATCAAAAGAAACTTATTAAATGAATTTGACAGGATAAT

cDNA AGAAAATCAAGAAAAATCCTTAAAGGCTTCAAAAAGCACTCCAGATG
genomic AAATcAaGAAAA-TCCTTAAAGGNTTCAAAAAGCACTCCAGATG

genomic gtaaaattagctttttatttatatctgttctccctctataggtatggtat

genomic ataatattctgacctcaggtgatccacctgcctctcaaagtgctgggatt

genomic acagacatgagccactgtgcctaatcaagggacctctttatactcttaaa

genomic aattactgaggacctaaaagagcatttggttatgtggaatatatctattg

Fig.2 (c).

exon29d

genomic tgtactgtgagttatttggtgcatagtcattatcaatttgtgaatcaatt

genomic tattttcatagttaacatttattgagcatccgttacattcactgaaaatt

genomic gtaaagcctataattgtctcaaatttttttgtgtatttacagtaacatgga

genomic tattctcttagattttaactaatatgtaatataaaataattgtttcctag

cDNA GCACAATAAAAGATCGAAGATTGTTTATGCATCATGTTTCTTTAGAGCCG
genomic GCACAATAAAAGATCGAAGATTGTAGGGCATCAGTT AGCCG

cDNA ATTACCTGTGTACCCTTTCG
genomic ATTACCTGTGTACCCTTTCGgtaagacatgtttaaatttttctaaattct

genomic aatacagtatgagaaaagtctcgtttttataaatgaacatttctaaaaat

genomic aatgacactaacgttaagaagttaacacttcccgttttataaaatttata

genomic aaatactttggtagtattttatagtgctgttcatatcattattttatttt

genomic ttaattttatgacagctttgtaaagtagacagattttattctaattttat

Fig.2(d).

exon bc

genomic acacactgagaggtccacacttgacagatttattntattatgngttatgt

genomic gaggtagattgtaagtcaaaggctagccttgaaaaatgtgatattgtttt

genomic ggaatggcaaccatggngaanacaaaacagtnaccagaatngtatcacca

genomic tgtagcaaatgagggtctgcaacaaaggcatattcataaatattnanang

cDNA
genomic tgtagtagtcaatnaacttataaattttntccccantgcagCACAACTAA
genomic CACNANTAA

cDNA GGAACGTCAAGAGATACAGAATCCAAATTTTACCGCACCTGGTCAAGAAT
genomic GGGACGTCAAGAGATACAGAATCCANATTTTACCGCACCTGGTCANGGAT

cDNA TTCTGTCTAAATCTCATTTGTATGAACATCTGACTTTGGAAAAATCTTCA
genomic TTCTGTNTAAATATCATTTGTATGAACATCTGACTTTGGAAAAATCTTCA

cDNA AGCAATTTAGCAGTTTCAGGACATCCATTTTATCAAGTTTCTGCTACAAG
genomic AGCAATTTNGCAGTTTCAGGACATCCATTTTATCAAGTTTCTGCTACAAG

cDNA AAATGAAAAAATGAGACACTTGATTACTACAGGCAGACCAACCAAAGTCT
genomic AAATGAAAAAATGAGACACTTGATTACTACAGGCAGACCAACCAAAGTCT

cDNA TTGTTCCACCTTTTAAAACTAAATCACATTTTC-ACAGAGTTGAACAGTGT
genomic TTGTTCCACCTTTTAAAACTAAATCACATTTTCCACAGAGTTGAACAGTGT

cDNA GTTAGGAATATTAACTTGGAGGAAAACAGACAAAAGCAAAACATTGATGG
genomic ggangaaacatATTNAAGCAAAACATTGATGG

cDNA ACATGGCTCTGATGATAGTAAAAATAAGATTAATGACAATGAGATTCATC
genomic ACNTGGCTCTGATGATAGTAAAAATAAGATTAATGACAATGAGATTCATC

cDNA AGTTTAACAGNAACAACTCCAATCAAGCAGCAGCTGTAACTTTCACAAAG
genomic AGTTTAACAAAAACAACTCCAATCAAGCAGCAGCTGTAACTTTCACAAAG

cDNA TGTGAAGAAGAACCTTTAG
genomic TGTGAANAAAAACCTTTAGgtattgtatgacaatttgtgtgatgaatttt

genomic tgcctttcagttanatatttccgttgttaaataatgtcctgatggtttnc

Fig.2 (e).

genomic ccccttggtggtggtaattttaaagccctttttaatgttttagattttc

genomic taaatccaaagattaggtttaaattattctaatgtttctttcaaanataa

exon a

cDNA ATTTAATTACAAGTCTTCAGAATGCCAGAGA

cDNA TATACAGGATATGCGAATTAAGAAGAAACAAAGGCAACGCGTCTTTCCAC

cDNA AGCCAGGCAGTCTGTATCTTGCAAAAACATCCACTCTGCCTCGAATCTCT

cDNA CTGAAAGCAGCAGTAGGAGGCCAAGTTCCCTCTGCGTGTTCTCATAAACA

cDNA GCTGTATACGTATGGCGTTTCTAAACATTGCATAAAAATTAACAGCAAAA

cDNA ATGCAGAGTCTTTTCAGTTTCACACTGAAGATTATTTTGGTAAGGAAAGT

cDNA TTATGGACTGGAAAAGGAATACAGTTGGCTGATGGTGGATGGCTCATACC

cDNA CTCCAATGATGGAAAGGCTGGAAAAGAAGAATTTTATAGGGCTCTGTGTG

cDNA ACGTAAAGGCCACATAGTGGATAAACCATGTGCCACTTGAGGAATTGAGA

cDNA AAAAGTCAGTGTGTTTAGAACATAAGGAGAGGGGCAGAGAAATGGATCTG

cDNA GTACAGGAGAATCAATGAGACCGGGTAAGAAACAGAAAAGGGGCTGCACC

cDNA AAATGATTGACACGCTCTGCAAACTTCTTTTGTTGGCTCAAGTTGTGGCT

cDNA CAAGAGGTGAGAAGGTAAGGCCAGACTTATTTGTTATGATTTGCCCTTTA

cDNA AATTGAAGCCTTAAGATTGGAATTCGATATCAAGCTTATCGATACCGTCG

cDNA ACCTCGAG

Fig.3.

BRCA2 protein sequence

Plaque 14 was translated in all 3 possible frames. The following protein sequence is the most likely of the 3 frames.

KVGSCLGGNLIFRXAXSXIHLRFRPKCIKNTSFVXIRETQSTVTPEMEKT
CSKEFKLSNNLNVEGGXSENNHSIKVSPYLSQFQQDKQQLVLGTKVSLVE
NIHVLGKEQASPKNVKMEIGKTETFSDVPVXTNIEVCSTYSKDSENYFET
EAVEIAKAFMEDDELTDXKLPSHATHXLFTCPENEEMVLSNSRIGKRRGE
PLILVGEPSIKRNLLNEFDRIIENQEKSLKASKSTPDGTIKDRLFVHHVS
LEPITCVPFRTTKERQEIQNPNFTAPGQEFLSKSHLYEHLTLEKSSSNLA
VSGHPFYQVSATRNEKMRHLITTGRPTKVFVPPFKTKSHFHRVEQCVRNI
NLEENRQKQNIDGHGSDDSKNKINDNEIHQFNKNNSNQAAAVTFTKCEEE
PLDLITSLQNARDIQDMRIKKKQRQRVFPQPGSLYLAKTSTLPRISLKAA
VGGQVPSACSHKQLYTYGVSKHCIKINSKNAESFQFHTEDYFGKESLWTG
KGIQLADGGWLIPSNDGKAGKEEFYRALCDVKAT

Fig.4(a).

BRCA2 cDNA

CCACATTGGAAAGTCAATGCCAAATGTCCTAGAAGATGAAGTATATGAAACAGTTGTAGA

TACCTCTGAAGAAGATAGTTTTTCATTATGTTTTTCTAAATGTAGAACAAAAAATCTACA

AAAAGTAAGAACTAGCAAGACTAGGAAAAAAATTTTCCATGAAGCAAACGCTGATGAATG

TGAAAAATCTAAAAACCAAGTGAAAGAAAAATACTCATTTGTATCTGAAGTGGAACCAAA

TGATACTGATCCATTAGATTCAAATGTAGCAAATCAGAAGCCCTTTGAGAGTGGAAGTGA

CAAAATCTCCAAGGAAGTTGTACCGTCTTTGGCCTGTGAATGGTCTCAACTAACCCTTTC

AGGTCTAAATGGAGCCCAGATGGAGAAAATACCCCTATTGCATATTTCTTCATGTGACCA

AAATATTTCAGAAAAAGACCTATTAGACACAGAGAACAAAAGAAAGAAAGATTTTCTTAC

TTCAGAGAATTCTTTGCCACGTATTTCTAGCCTACCAAAATCAGAGAAGCCATTAAATGA

GGAAACAGTGGTAAATAAGAGAGATGAAGAGCAGCATCTTGAATCTCATACAGACTGCAT

TCTTGCAGTAAAGCAGGCAATATCTGGAACTTCTCCAGTGGCTTCTTCATTTCAGGGTAT

CAAAAAGTCTATATTCAGAATAAGAGAATCACCTAAAGAGACTTTCAATGCAAGTTTTTC

AGGTCATATGACTGATCCAAACTTTAAAAAAGAAACTGAAGCCTCTGAAAGTGGACTGGA

AATACATACTGTTTGCTCACAGAAGGAGGACTCCTTATGTCCAAATTTAATTGATAATGG

AAGCTGGCCAGCCACCACCACACAGAATTCTGTAGCTTTGAAGAATGCAGGTTTAATATC

CACTTTGAAAAAGAAAACAAATAAGTTTATTTATGCTATACATGATGAAACATCTTATAA

AGGAAAAAAAATACCGAAAGACCAAAAATCAGAACTAATTAACTGTTCAGCCCAGTTTGA

AGCAAATGCTTTTGAAGCACCACTTACATTTGCAAATGCTGATTCAGGTTTATTGCATTC

TTCTGTGAAAAGAAGCTGTTCACAGAATGATTCTGAAGAACCAACTTTGTCCTTAACTAG

CTCTTTTGGGACAATTCTGAGGAAATGTTCTAGAAATGAAACATGTTCTAATAATACAGT

AATCTCTCAGGATCTTGATTATAAAGAAGCAAAATGTAATAAGGAAAAACTACAGTTATT

TATTACCCCAGAAGCTGATTCTCTGTCATGCCTGCAGGAAGGACAGTGTGAAAATGATCC

AAAAAGCAAAAAAGTTTCAGATATAAAAGAAGAGGTCTTGGCTGCAGCATGTCACCCAGT

ACAACATTCAAAAGTGGAATACAGTGATACTGACTTTCAATCCCAGAAAAGTCTTTTATA

TGATCATGAAAATGCCAGCACTCTTATTTTAACTCCTACTTCCAAGGATGTTCTGTCAAA

CCTAGTCATGATTTCTAGAGGCAAAGAATCATACAAAATGTCAGACAAGCTCAAAGGTAA

CAATTATGAATCTGATGTTGAATTAACCAAAAATATTCCCATGGAAAAGAATCAAGATGT

ATGTGCTTTAAATGAAAATTATAAAAACGTTGAGCTGTTGCCACCTGAAAAATACATGAG

Fig.4(b).

AGTAGCATCACCTTCAAGAAAGGTACAATTCAACCAAAACACAAATCTAAGAGTAATCCA

AAAAAATCAAGAAGAAACTACTTCAATTTCAAAAATAACTGTCAATCCAGACTCTGAAGA

ACTTTTCTCAGACAATGAGAATAATTTTGTCTTCCAAGTAGCTAATGAAAGGAATAATCT

TGCTTTAGGAAATACTAAGGAACTTCATGAAACAGACTTGACTTGTGTAAACGAACCCAT

TTTCAAGAACTCTACCATGGTTTTATATGGAGACACAGGTGATAAACAAGCAACCCAAGT

GTCAATTAAAAAAGATTTGGTTTATGTTCTTGCAGAGGAGAACAAAAATAGTGTAAAGCA

GCATATAAAAATGACTCTAGGTCAAGATTTAAAATCGGACATCTCCTTGAATATAGATAA

AATACCAGAAAAAAATAATGATTACATGAACAAATGGGCAGGACTCTTAGGTCCAATTTC

AAATCACAGTTTTGGAGGTAGCTTCAGAACAGCTTCAAATAAGGAAATCAAGCTCTCTGA

ACATAACATTAAGAAGAGCAAAATGTTCTTCAAAGATATTGAAGAACAATATCCTACTAG

TTTAGCTTGTGTTGAAATTGTAAATACCTTGGCATTAGATAATCAAAAGAAACTGAGCAA

GCCTCAGTCAATTAATACTGTATCTGCACATTTACAGAGTAGTGTAGTTGTTTCTGATTG

TAAAAATAGTCATATAACCCCTCAGATGTTATTTTCCAAGCAGGATTTTAATTCAAACCA

TAATTTAACACCTAGCCAAAAGGAGCAAATTACAGAACTTTCTACTATATTAGAAGATTC

AGGAAGTCAGTTTGAATTTACTCAGTTTAGAAAACCAAGCTACATATTGCAGAAGAGTAC

ATTTGAAGTGCCTGAAAACCAGATGACTATCTTAAAGACCACTTCTGAGGAATGCAGAGA

TGCTGATCTTCATGTCATAATGAATGCCCCATCGATTGGTCAGGTAGACAGCAGCAAGCA

ATTTGAAGGTACAGTTGAAATTAAACGGAAGTTTGCTGGCCTGTTGAAAAATGACTGTAA

CAAAAGTGCTTCTGGTTATTTAACAGATGAAAATGAAGTGGGGTTTAGGGGCTTTTATTC

TGCTCATGGCACAAAACTGAATGTTTCTACTGAAGCTCTGCAAAAAGCTGTGAAACTGTT

TAGTGATATTGAGAATATTAGTGAGGAAACTTCTGCAGAGGTACATCCAATAAGTTTATC

TTCAAGTAAATGTCATGATTCTGTCGTTTCAATGTTTAAGATAGAAAATCATAATGATAA

AACTGTAAGTGAAAAAAATAATAAATGCCAACTGATATTACAAAATAATATTGAAATGAC

TACTGGCACTTTTGTTGAAGAAATTACTGAAAATTACAAGAGAAATACTGAAAATGAAGA

TAACAAATATACTGCTGCCAGTAGAAATTCTCATAACTTAGAATTTGATGGCAGTGATTC

AAGTAAAAATGATACTGTTTGTATTCATAAAGATGAAACGGACTTGCTATTTACTGATCA

GCACAACATATGTCTTAAATTATCTGGCCAGTTTATGAAGGAGGGAAACACTCAGATTAA

AGAAGATTTGTCAGATTTAACTTTTTTGGAAGTTGCGAAAGCTCAAGAAGCATGTCATGG

TAATACTTCAAATAAAGAACAGTTAACTGCTACTAAAACGGAGCAAAATATAAAAGATTT

Fig.4(c).

TGAGACTTCTGATACATTTTTTCAGACTGCAAGTGGGAAAAATATTAGTGTCGCCAAAGA

GTCATTTAATAAAATTGTAAATTTCTTTGATCAGAAACCAGAAGAATTGCATAACTTTTC

CTTAAATTCTGAATTACATTCTGACATAAGAAAGAACAAAATGGACATTCTAAGTTATGA

GGAAACAGACATAGTTAAACACAAAATACTGAAAGAAAGTGTCCCAGTTGGTACTGGAAA

TCAACTAGTGACCTTCCAGGGACAACCCGAACGTGATGAAAAGATCAAAGAACCTACTCT

GTTGGGTTTTCATACAGCTAGCGGGAAAAAAGTTAAAATTGCAAAGGAATCTTTGGACAA

AGTGAAAAACCTTTTTGATGAAAGAGCAAGGTACTAGTGAAATCACCAGTTTTAGCCATC

AATGGGCAAAGACCCTAAAGTACAGAGAGGCCTGTAAAGACCTTGAATTAGCATGTGAGA

CCATTGAGATCACAGCTGCCCCAAAGTGTAAAGAAATGCAGAATTCTCTCAATAATGATA

AAAACCTTGTTTCTATTGAGACTGTGGTGCCACCTAAGCTCTTAAGTGATAATTTATGTA

GACAAACTGAAAATCTCAAAACATCAAAAAGTATCTTTTTGAAAGTTAAAGTACATGAAA

ATGTAGAAAAAGAAACAGCAAAAAGTCCTGCAACTTGTTACACAAATCAGTCCCCTTATT

CAGTCATTGAAAATTCAGCCTTAGCTTTTTACACAAGTTGTAGTAGAAAAACTTCTGTGA

GTCAGACTTCATTACTTGAAGCAAAAAAATGGCTTAGAGAAGGAATATTTGATGGTCAAC

CAGAAAGAATAAATACTGCAGATTATGTAGGAAATTATTTGTATGAAAATAATTCAAACA

GTACTATAGCTGAAAATGACAAAAATCATCTCTCCGAAAAACAAGATACTTATTTAAGTA

ACAGTAGCATGTCTAACAGCTATTCCTACCATTCTGATGAGGTATATAATGATTCAGGAT

ATCTCTCAAAAAATAAACTTGATTCTGGTATTGAGCCAGTATTGAAGAATGTTGAAGATC

AAAAAAACACTAGTTTTTCCAAAGTAATATCCAATGTAAAAGATGCAAATGCATACCCAC

AAACTGTAAATGAAGATATTTGCGTTGAGGAACTTGTGACTAGCTCTTCACCCTGCAAAA

ATAAAAATGCAGCCATTAAATTGTCCATATCTAATAGTAATAATTTTGAGGTAGGGCCAC

CTGCATTTAGGATAGCCAGTGGTAAAATCGTTTGTGTTTCACATGAAACAATTAAAAAAG

TGAAAGACATATTTACAGACAGTTTCAGTAAAGTAATTAAGGAAAACAACGAGAATAAAT

CAAAAATTTGCCAAACGAAAATTATGGCAGGTTGTTACGAGGCATTGGATGATTCAGAGG

ATATTCTTCATAACTCTCTAGATAATGATGAATGTAGCACGCATTCACATAAGGTTTTTG

CTGACATTCAGAGTGAAGAAATTTTACAACATAACCAAAATATGTCTGGATTGGAGAAAG

TTTCTAAAATATCACCTTGTGATGTTAGTTTGGAAACTTCAGATATATGTAAATGTAGTA

TAGGGAAGCTTCATAAGTCAGTCTCATCTGCAAATACTTGTGGGATTTTTAGCACAGCAA

GTGGAAAATCTGTCCAGGTATCAGATGCTTCATTACAAAACGCAAGACAAGTGTTTTCTG

Fig.4(d).

AAATAGAAGATAGTACCAAGCAAGTCTTTTCCAAAGTATTGTTTAAAAGTAACGAACATT

CAGACCAGCTCACAAGAGAAGAAAATACTGCTATACGTACTCCAGAACATTTAATATCCC

AAAAAGGCTTTTCATATAATGTGGTAAATTCATCTGCTTTCTCTGGATTTAGTACAGCAA

GTGGAAAGCAAGTTTCCATTTTAGAAAGTTCCTTACACAAAGTTAAGGGAGTGTTAGAGG

AATTTGATTTAATCAGAACTGAGCATAGTCTTCACTATTCACCTACGTCTAGACAAAATG

TATCAAAAATACTTCCTCGTGTTGATAAGAGAAACCCAGAGCACTGTGTAAACTCAGAAA

TGGAAAAAACCTGCAGTAAAGAATTTAAATTATCAAATAACTTAAATGTTGAAGGTGGTT

CTTCAGAAAATAATCACTCTATTAAAGTTTCTCCATATCTCTCTCAATTTCAACAAGACA

AACAACAGTTGGTATTAGGAACCAAAGTCTCACTTGTTGAGAACATTCATGTTTTGGGAA

AAGAACAGGCTTCACCTAAAAACGTAAAAATGGAAATTGGTAAAACTGAAACTTTTTCTG

ATGTTCCTGTGAAAACAAATATAGAAGTTTGTTCTACTTACTCCAAAGATTCAGAAAACT

ACTTTGAAACAGAAGCAGTAGAAATTGCTAAAGCTTTTATGGAAGATGATGAACTGACAG

ATTCTAAACTGCCAAGTCATGCCACACATTCTCTTTTTACATGTCCCGAAAATGAGGAAA

TGGTTTTGTCAAATTCAAGAATTGGAAAAAGAAGAGGAGAGCCCCTTATCTTAGTGGGAG

AACCCTCAATCAAAAGAAACTTATTAAATGAATTTGACAGGATAATAGAAAATCAAGAAA

AATCCTTAAAGGCTTCAAAAAGCACTCCAGATGGCACAATAAAAGATCGAAGATTGTTTG

TGCATCATGTTTCTTTAGAGCCGATTACCTGTGTACCCTTTCGCACAACTAAGGAACGTC

AAGAGATACAGAATCCAAATTTTACCGCACCTGGTCAAGAATTTCTGTCTAAATCTCATT

TGTATGAACATCTGACTTTGGAAAAATCTTCAAGCAATTTAGCAGTTTCAGGACATCCAT

TTTATCAAGTTTCTGGTAACAAGAATGGAAAAATGAGAAAATTGATTACTACAGGCAGAC

CAACCAAAGTCTTTGTTCCACCTTTTAAAACTAAATCACATTTTCACAGAGTTGAACAGT

GTGTTAGGAATATTAACTTGGAGGGAAACAGACAAAAGCAAAACATTGATGGACATGGCT

CTGATGATAGTAAAAATAAGATTAATGACAATGAGATTCATCAGTTTAACAAAAACAACT

CCAATCAAGCAGCAGCTGTAACTTTCACAAAGTGTGAAGAAGAACCTTTAGATTTAATTA

CAAGTCTTCAGAATGCCAGAGATATACAGGATATGCGAATTAAGAAGAAACAAAGGCAAC

GCGTCTTTCCACAGCCAGGCAGTCTGTATCTTGCAAAAACATCCACTCTGCCTCGAATCT

CTCTGAAAGCAGCAGTAGGAGGCCAAGTTCCCTCTGCGTGTTCTCATAAACAGCTGTATA

CGTATGGCGTTTCTAAACATTGCATAAAAATTAACAGCAAAAATGCAGAGTCTTTTCAGT

TTCACACTGAAGATTATTTTGGTAAGGAAAGTTTATGGACTGGAAAAGGAATACAGTTGG

Fig.4(e).

CTGATGGTGGATGGCTCATACCCTCCAATGATGGAAAGGCTGGAAAAGAAGAATTTTATA

GGGCTCTGTGTGACGTAAAGGCCACATAGTGGATAAACCATGTGCCACTTGAGGAATTGA

GAAAAAGTCAGTGTGTTTAGAACATAAGGAGAGGGGCAGAGAAATGGATCTGGTACAGGA

GAATCAATGAGACCGGGTAAGAAACAGAAAAGGGGCTGCACCAAATGATTGACACGCTCT

GCAAACTTCTTTTGTTGGCTCAAGTTGTGGCTCAAGAGGTGAGAAGGTAAGGCCAGACTT

ATTTGTTATGATTTGCCCTTTAAATTGAAGCCTTAAGATT

Fig.5(a).

HisIleGlyLysSerMetProAsnValLeuGluAspGluValTyrGluThrValValAsp 20
ThrSerGluGluAspSerPheSerLeuCysPheSerLysCysArgThrLysAsnLeuGln 40
LysValArgThrSerLysThrArgLysLysIlePheHisGluAlaAsnAlaAspGluCys 60
GluLysSerLysAsnGlnValLysGluLysTyrSerPheValSerGluValGluProAsn 80
AspThrAspProLeuAspSerAsnValAlaAsnGlnLysProPheGluSerGlySerAsp 100
LysIleSerLysGluValValProSerLeuAlaCysGluTrpSerGlnLeuThrLeuSer 120
GlyLeuAsnGlyAlaGlnMetGluLysIleProLeuLeuHisIleSerSerCysAspGln 140
AsnIleSerGluLysAspLeuLeuAspThrGluAsnLysArgLysLysAspPheLeuThr 160
SerGluAsnSerLeuProArgIleSerSerLeuProLysSerGluLysProLeuAsnGlu 180
GluThrValValAsnLysArgAspGluGluGlnHisLeuGluSerHisThrAspCysIle 200
LeuAlaValLysGlnAlaIleSerGlyThrSerProValAlaSerSerPheGlnGlyIle 220
LysLysSerIlePheArgIleArgGluSerProLysGluThrPheAsnAlaSerPheSer 240
GlyHisMetThrAspProAsnPheLysLysGluThrGluAlaSerGluSerGlyLeuGlu 260
IleHisThrValCysSerGlnLysGluAspSerLeuCysProAsnLeuIleAspAsnGly 280
SerTrpProAlaThrThrThrGlnAsnSerValAlaLeuLysAsnAlaGlyLeuIleSer 300
ThrLeuLysLysLysThrAsnLysPheIleTyrAlaIleHisAspGluThrSerTyrLys 320
GlyLysLysIleProLysAspGlnLysSerGluLeuIleAsnCysSerAlaGlnPheGlu 340
AlaAsnAlaPheGluAlaProLeuThrPheAlaAsnAlaAspSerGlyLeuLeuHisSer 360
SerValLysArgSerCysSerGlnAsnAspSerGluGluProThrLeuSerLeuThrSer 380
SerPheGlyThrIleLeuArgLysCysSerArgAsnGluThrCysSerAsnAsnThrVal 400
IleSerGlnAspLeuAspTyrLysGluAlaLysCysAsnLysGluLysLeuGlnLeuPhe 420
IleThrProGluAlaAspSerLeuSerCysLeuGlnGluGlyGlnCysGluAsnAspPro 440
LysSerLysLysValSerAspIleLysGluGluValLeuAlaAlaAlaCysHisProVal 460
GlnHisSerLysValGluTyrSerAspThrAspPheGlnSerGlnLysSerLeuLeuTyr 480
AspHisGluAsnAlaSerThrLeuIleLeuThrProThrSerLysAspValLeuSerAsn 500
LeuValMetIleSerArgGlyLysGluSerTyrLysMetSerAspLysLeuLysGlyAsn 520
AsnTyrGluSerAspValGluLeuThrLysAsnIleProMetGluLysAsnGlnAspVal 540
CysAlaLeuAsnGluAsnTyrLysAsnValGluLeuLeuProProGluLysTyrMetArg 560
ValAlaSerProSerArgLysValGlnPheAsnGlnAsnThrAsnLeuArgValIleGln 580
LysAsnGlnGluGluThrThrSerIleSerLysIleThrValAsnProAspSerGluGlu 600
LeuPheSerAspAsnGluAsnAsnPheValPheGlnValAlaAsnGluArgAsnAsnLeu 620
AlaLeuGlyAsnThrLysGluLeuHisGluThrAspLeuThrCysValAsnGluProIle 640
PheLysAsnSerThrMetValLeuTyrGlyAspThrGlyAspLysGlnAlaThrGlnVal 660
SerIleLysLysAspLeuValTyrValLeuAlaGluGluAsnLysAsnSerValLysGln 680
HisIleLysMetThrLeuGlyGlnAspLeuLysSerAspIleSerLeuAsnIleAspLys 700
IleProGluLysAsnAsnAspTyrMetAsnLysTrpAlaGlyLeuLeuGlyProIleSer 720
AsnHisSerPheGlyGlySerPheArgThrAlaSerAsnLysGluIleLysLeuSerGlu 740
HisAsnIleLysLysSerLysMetPhePheLysAspIleGluGluGlnTyrProThrSer 760

Fig.5(b).

LeuAlaCysValGluIleValAsnThrLeuAlaLeuAspAsnGlnLysLysLeuSerLys 780
ProGlnSerIleAsnThrValSerAlaHisLeuGlnSerSerValValValSerAspCys 800
LysAsnSerHisIleThrProGlnMetLeuPheSerLysGlnAspPheAsnSerAsnHis 820
AsnLeuThrProSerGlnLysGluGlnIleThrGluLeuSerThrIleLeuGluAspSer 840
GlySerGlnPheGluPheThrGlnPheArgLysProSerTyrIleLeuGlnLysSerThr 860
PheGluValProGluAsnGlnMetThrIleLeuLysThrThrSerGluGluCysArgAsp 880
AlaAspLeuHisValIleMetAsnAlaProSerIleGlyGlnValAspSerSerLysGln 900
PheGluGlyThrValGluIleLysArgLysPheAlaGlyLeuLeuLysAsnAspCysAsn 920
LysSerAlaSerGlyTyrLeuThrAspGluAsnGluValGlyPheArgGlyPheTyrSer 940
AlaHisGlyThrLysLeuAsnValSerThrGluAlaLeuGlnLysAlaValLysLeuPhe 960
SerAspIleGluAsnIleSerGluGluThrSerAlaGluValHisProIleSerLeuSer 980
SerSerLysCysHisAspSerValValSerMetPheLysIleGluAsnHisAsnAspLys 1000
ThrValSerGluLysAsnAsnLysCysGlnLeuIleLeuGlnAsnAsnIleGluMetThr 1020
ThrGlyThrPheValGluGluIleThrGluAsnTyrLysArgAsnThrGluAsnGluAsp 1040
AsnLysTyrThrAlaAlaSerArgAsnSerHisAsnLeuGluPheAspGlySerAspSer 1060
SerLysAsnAspThrValCysIleHisLysAspGluThrAspLeuLeuPheThrAspGln 1080
HisAsnIleCysLeuLysLeuSerGlyGlnPheMetLysGluGlyAsnThrGlnIleLys 1100
GluAspLeuSerAspLeuThrPheLeuGluValAlaLysAlaGlnGluAlaCysHisGly 1120
AsnThrSerAsnLysGluGlnLeuThrAlaThrLysThrGluGlnAsnIleLysAspPhe 1140
GluThrSerAspThrPhePheGlnThrAlaSerGlyLysAsnIleSerValAlaLysGlu 1160
SerPheAsnLysIleValAsnPhePheAspGlnLysProGluGluLeuHisAsnPheSer 1180
LeuAsnSerGluLeuHisSerAspIleArgLysAsnLysMetAspIleLeuSerTyrGlu 1200
GluThrAspIleValLysHisLysIleLeuLysGluSerValProValGlyThrGlyAsn 1220
GlnLeuValThrPheGlnGlyGlnProGluArgAspGluLysIleLysGluProThrLeu 1240
LeuGlyPheHisThrAlaSerGlyLysLysValLysIleAlaLysGluSerLeuAspLys 1260
ValLysAsnLeuPheAspGluLysGluGlnGlyThrSerGluIleThrSerPheSerHis 1280
GlnTrpAlaLysThrLeuLysTyrArgGluAlaCysLysAspLeuGluLeuAlaCysGlu 1300
ThrIleGluIleThrAlaAlaProLysCysLysGluMetGlnAsnSerLeuAsnAsnAsp 1320
LysAsnLeuValSerIleGluThrValValProProLysLeuLeuSerAspAsnLeuCys 1340
ArgGlnThrGluAsnLeuLysThrSerLysSerIlePheLeuLysValLysValHisGlu 1360
AsnValGluLysGluThrAlaLysSerProAlaThrCysTyrThrAsnGlnSerProTyr 1380
SerValIleGluAsnSerAlaLeuAlaPheTyrThrSerCysSerArgLysThrSerVal 1400
SerGlnThrSerLeuLeuGluAlaLysLysTrpLeuArgGluGlyIlePheAspGlyGln 1420
ProGluArgIleAsnThrAlaAspTyrValGlyAsnTyrLeuTyrGluAsnAsnSerAsn 1440
SerThrIleAlaGluAsnAspLysAsnHisLeuSerGluLysGlnAspThrTyrLeuSer 1460
AsnSerSerMetSerAsnSerTyrSerTyrHisSerAspGluValTyrAsnAspSerGly 1480
TyrLeuSerLysAsnLysLeuAspSerGlyIleGluProValLeuLysAsnValGluAsp 1500
GlnLysAsnThrSerPheSerLysValIleSerAsnValLysAspAlaAsnAlaTyrPro 1520

Fig.5(c).

GlnThrValAsnGluAspIleCysValGluGluLeuValThrSerSerSerProCysLys 1540
AsnLysAsnAlaAlaIleLysLeuSerIleSerAsnSerAsnAsnPheGluValGlyPro 1560
ProAlaPheArgIleAlaSerGlyLysIleValCysValSerHisGluThrIleLysLys 1580
ValLysAspIlePheThrAspSerPheSerLysValIleLysGluAsnAsnGluAsnLys 1600
SerLysIleCysGlnThrLysIleMetAlaGlyCysTyrGluAlaLeuAspAspSerGlu 1620
AspIleLeuHisAsnSerLeuAspAsnAspGluCysSerThrHisSerHisLysValPhe 1640
AlaAspIleGlnSerGluGluIleLeuGlnHisAsnGlnAsnMetSerGlyLeuGluLys 1660
ValSerLysIleSerProCysAspValSerLeuGluThrSerAspIleCysLysCysSer 1680
IleGlyLysLeuHisLysSerValSerSerAlaAsnThrCysGlyIlePheSerThrAla 1700
SerGlyLysSerValGlnValSerAspAlaSerLeuGlnAsnAlaArgGlnValPheSer 1720
GluIleGluAspSerThrLysGlnValPheSerLysValLeuPheLysSerAsnGluHis 1740
SerAspGlnLeuThrArgGluGluAsnThrAlaIleArgThrProGluHisLeuIleSer 1760
GlnLysGlyPheSerTyrAsnValValAsnSerSerAlaPheSerGlyPheSerThrAla 1780
SerGlyLysGlnValSerIleLeuGluSerSerLeuHisLysValLysGlyValLeuGlu 1800
GluPheAspLeuIleArgThrGluHisSerLeuHisTyrSerProThrSerArgGlnAsn 1820
ValSerLysIleLeuProArgValAspLysArgAsnProGluHisCysValAsnSerGlu 1840
MetGluLysThrCysSerLysGluPheLysLeuSerAsnAsnLeuAsnValGluGlyGly 1860
SerSerGluAsnAsnHisSerIleLysValSerProTyrLeuSerGlnPheGlnGlnAsp 1880
LysGlnGlnLeuValLeuGlyThrLysValSerLeuValGluAsnIleHisValLeuGly 1900
LysGluGlnAlaSerProLysAsnValLysMetGluIleGlyLysThrGluThrPheSer 1920
AspValProValLysThrAsnIleGluValCysSerThrTyrSerLysAspSerGluAsn 1940
TyrPheGluThrGluAlaValGluIleAlaLysAlaPheMetGluAspAspGluLeuThr 1960
AspSerLysLeuProSerHisAlaThrHisSerLeuPheThrCysProGluAsnGluGlu 1980
MetValLeuSerAsnSerArgIleGlyLysArgArgGlyGluProLeuIleLeuValGly 2000
GluProSerIleLysArgAsnLeuLeuAsnGluPheAspArgIleIleGluAsnGlnGlu 2020
LysSerLeuLysAlaSerLysSerThrProAspGlyThrIleLysAspArgArgLeuPhe 2040
ValHisHisValSerLeuGluProIleThrCysValProPheArgThrThrLysGluArg 2060
GlnGluIleGlnAsnProAsnPheThrAlaProGlyGlnGluPheLeuSerLysSerHis 2080
LeuTyrGluHisLeuThrLeuGluLysSerSerSerAsnLeuAlaValSerGlyHisPro 2100
PheTyrGlnValSerGlyAsnLysAsnGlyLysMetArgLysLeuIleThrThrGlyArg 2120
ProThrLysValPheValProProPheLysThrLysSerHisPheHisArgValGluGln 2140
CysValArgAsnIleAsnLeuGluGlyAsnArgGlnLysGlnAsnIleAspGlyHisGly 2160
SerAspAspSerLysAsnLysIleAsnAspAsnGluIleHisGlnPheAsnLysAsnAsn 2180
SerAsnGlnAlaAlaAlaValThrPheThrLysCysGluGluGluProLeuAspLeuIle 2200
ThrSerLeuGlnAsnAlaArgAspIleGlnAspMetArgIleLysLysLysGlnArgGln 2220
ArgValPheProGlnProGlySerLeuTyrLeuAlaLysThrSerThrLeuProArgIle 2240
SerLeuLysAlaAlaValGlyGlyGlnValProSerAlaCysSerHisLysGlnLeuTyr 2260
ThrTyrGlyValSerLysHisCysIleLysIleAsnSerLysAsnAlaGluSerPheGln 2280

Fig.5(d).

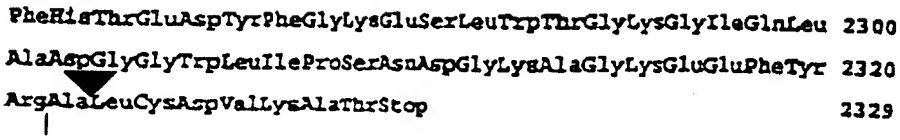

PheHisThrGluAspTyrPheGlyLysGluSerLeuTrpThrGlyLysGlyIleGlnLeu 2300
AlaAspGlyGlyTrpLeuIleProSerAsnAspGlyLysAlaGlyLysGluGluPheTyr 2320
ArgAlaLeuCysAspValLysAlaThrStop 2329

Fig.6.

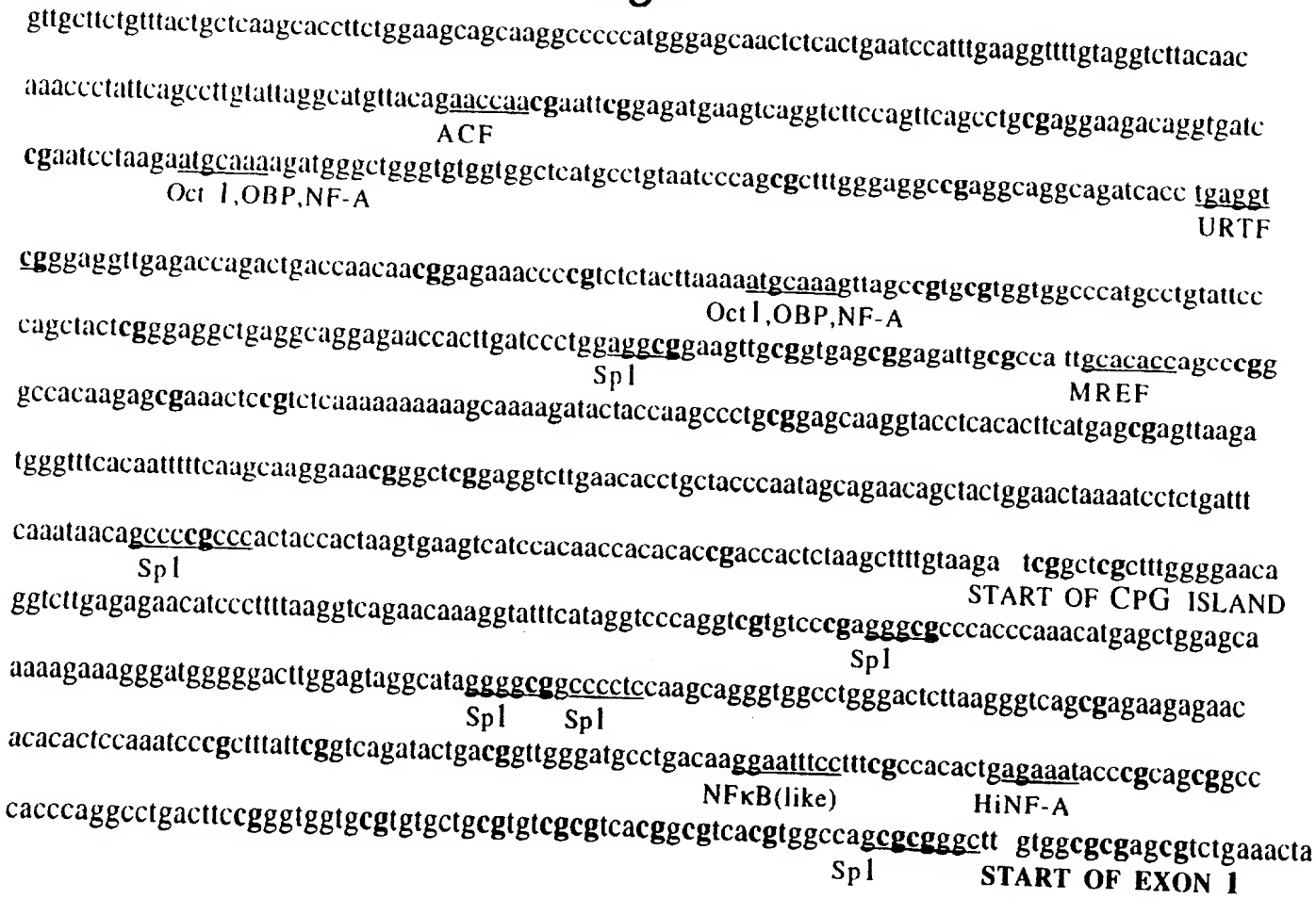

ACF
Oct 1,OBP,NF-A
URTF
Oct 1,OBP,NF-A
Sp1
MREF
Sp1
START OF CPG ISLAND
Sp1
Sp1 Sp1
NFκB(like)
HiNF-A
Sp1
START OF EXON 1

Fig.6 (Cont).

gcggcagagcggagccgctgtggcactctgcgcctctctgcgcctcgggtgtcttttcggcggtgggtcgccgccgggagaagcgtgaggggac

agatttgtgaccggcgcggtttttgtcagcttactccggccaaaaaagaactgcgcctctggagcgggttagtggtggtggtagtgggttgggacga

gcgcgtcttccgcagtcccagtccagcgtggcgggggagcgcctcacgcccgggtcgctgccgcggcttcttgccctttgtctctgccaaccccc

acccatgcctgagagaaaggtccttgccgaaggcaaattttcgccaagcaaattcgagccccgccccttccctgggtctccatttccgcctccgg

Sp1

ccggccttttgggctccgccttcagctcaagacttaacttccctccccagctgtcccagatgacgccatctgaaatttcttggaaacacgatcactttaac

Sp1 / RiPE3B LyF-1 IEF-1, LEF1

ggaatattgctgttttggggaagtgttttacagctgctgggcacgctgtatttgccttacttaagcccctggtaattgctgtattccgaagacatgctgatg

ggaattaccaggcggcgttggtctctaaactgagccctctgtcccaactagccacgcgtcactggttagcg tgattgaaactaaatcgtatgaaaatc

CP2 END OF CPG ISLAND

ctcttctctagtcgcactagccacgtttcgagtgcttaatgtggctagtggcaccggtttggacagcacagctgtaaaatgttcccatcctcacagtaag

CP2

ctgttaccgttccaggagatgggactgaattagaattcaaacaaatttccagcgcttctgagttttacctcagtcacataataaggaatgcatccctgtg

taagtgcattttggtcttctgttttgcagacttatttaccaagcattggaggaatatcgtaggtaaaaatgcctattggatccaaagagaggccaacatttt

ttgaaatttttaagacacgctgcaacaaagcaggtattgac

Putative *BRCA2* promoter region. Potential binding sites for various transcription factors are underlined. CpG are emboldened.

Fig.7. (1/22)

The exon/intron structure of the BRCA2 gene.

the genomic sequence available from:
ftp://sanger.ac.uk/pub/human/sequences/13q
and ftp://genome.wustl.edu/pub/gsc1/brca2
and is present in the Pacs 214k23 and 92m18

Exon	1	2	3	4	5	6	7	8	9
Size	188	106	249	109	50	41	115	50	112

Exon	10	11	12	13	14	15	16	17	18
Size	1116	4932	96	70	428	166	204	171	355

Exon	19	20	21	22	23	24	25	26	27
Size	156	145	122	199	164	139	245	147	606

NB there is 67bp of coding sequence in exon 2

Exon 1

.....aggggggncccccttcccaaaagnaag
ggtggcccttggnaacntcctnaaaggntcaangcnaantaaagaaagaaacacnaacaa
ctcccaaatccccgctttaattcggtcaagaatactaacggttgggaatgccttgaacaa
aggaaatttccttcgccaacactgagaaaatacccgcaagcgcccnacccnaggcctga
cttccgggtggtgcgtgtgctgcgtgtcgcgtcacggcgtcacgtggccagcgcgggctt
GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCGCTGTGGCACTGCTGCGCCT
CTGCTGCGCCTCGGGTGTCTTTTGCGGCGGTGGGTCGCCGCCGGGAGAAGCGTGAGGGGA
CAGATTTGTGACCGGCGCGGTTTTTGTCAGCTTACTCCGGCCAAAAAAGAACTGCGCCTC
TGGAGCGGgttagtggtggtggtagtgggttgggacgagcgcgtcttccgcagtcccagt
ccagcgtggcgggggagcgcctcacgccccgggtcgctgccgcggcttcttgccctttg
tctctgccaacccccacccatgcctgagagaaaggtccttgcccgaaggcaaattttcgc
caagcaaattcgagccccgcccctttccctgggtctccatttcccgcctccggcccggcct
ttgggctccgccttcagctcaagacttaacttccctcccagctgtcccagatgacgccat
ctgaaatttcttggaaacacgatcactttaacggaatattgctgttttggggaagtgttt
tacagctgctgggcacgctgtatttgccttacttaagcccctggtaattgctgtattccg
aagacatgctgatgggaattaccaggcggcgttggtctctaaactgagccctctgtccca
actagctacgcgtcactggttagcgtga.....

Exon 2

.gacttaacttcccctcccagctgtcccagatgacgccatctgaaatttcttggaaacacg
atcactttaacggaatattgctgttttggggaagtgttttacagctgctgggcacgctgt
atttgccttacttaagcccctggtaattgctgtattccgaagacatgctgatgggaatta
ccaggcggcgttggtctctaaactgagccctctgtcccaactagctacgcgtcactggtt

Fig.7. (2/22)

```
agcgtgattgaaactaaatcgtatgaaaatcctcttctctagtcgcactagccacgtttc
gagtgcttaatgtggctagtggcaccggtttggacagcacagctgtaaaatgttcccatc
ctcacagtaagctgttaccgttccaggagatgggactgaattagaattcaaacaaatttt
ccagcgcttctgagttttacctcagtcacataataaggaatgcatccctgtgtaagtgca
ttttggtcttctgttttgcagACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAAAA

ATGCCTATTGGATCCAAAGAGAGGCCAACATTTTTTGAAATTTTTAAGACACGCTGCAAC
M  P  I  G  S  K  E  R  P  T  F  F  E  I  F  K  T  R  C  N     20

AAAGCAGgtattgacaaattttatataactttataaattacaccgagaaagtgttttcta
K  A  D                                                        23

aaaaatgcttgctaaaaacccagtacgtcacagtgttgcttagaaccataaactgttcct
tatgtgtgtataaatccagttaacaacataatcatcgtttgcaggttaaccacatgataa
atatagaacgtctagtggataaagaggaaactggccccttgactagcagtaggtaacaat
tactaacaaatcagaagcattaatgttactttatggcagaagttgtccaacttttggtt
tcagtactccttatactcttaaaaatgatctaggacccccggagtgcttttgtttatgta
gcttaccatattagaaatttaaaactaagaatttaaggctgggcgtggtggctcacgcct
gtaatcccagcactttgggaggcccgaggtgggcggatcacttgaggccagaagtttgag
accagcctggccaacatggtgaaaccc.....

Exon 3
.....gactacaggcgtgtgccacc
acgcttggctaatttttgtgtttttagtaaagatggggtttcaacgtgttagcaaggtt
ggtctcgatctgacctcgtgatctgctcgcctcagcctcccaaagtgttgggattacagg
cgtgagcccccgcacctggccgaattttatcgtggaatgtattcttaatgtgaatagttt
ttgattccgaaccatgaataataagaaaataaataaaatttaaatgaaaataaaagctaa
tatatacagcttttaataatatagttaaatgccatcttgtaacttttgtgaactcttgtt
acacctttctatagattcgcaagagaatggattaatgatcttgtttaattaatatgcctt
aacaaaagtaatccatagtcaagatcttaagcatttttttccttatgatctttaactgtt
ctgggtcacaaatttgtctgtcactggttaaaactaaggtgggatttttttttaaatag

ATTTAGGACCAATAAGTCTTAATTGGTTTGAAGAACTTTCTTCAGAAGCTCCACCCTATA
  L  G  P  I  S  L  N  W  F  E  E  L  S  S  E  A  P  P  Y  N  43

ATTCTGAACCTGCAGAAGAATCTGAACATAAAAACAACAATTACGAACCAAACCTATTTA
  S  E  P  A  E  E  S  E  H  K  N  N  N  Y  E  P  N  L  F  K  63

AAACTCCACAAAGGAAACCATCTTATAATCAGCTGGCTTCAACTCCAATAATATTCAAAG
  T  P  Q  R  K  P  S  Y  N  Q  L  A  S  T  P  I  I  F  K  E  83

AGCAAGGGCTGACTCTGCCGCTGTACCAATCTCCTGTAAAAGAATTAGATAAATTCAAAT
  Q  G  L  T  L  P  L  Y  Q  S  P  V  K  E  L  D  K  F  K  L 103

TAGACTTAGgtaagtaatgcaatatggtagactggggagaactacaaactaggaatttag
  D  L  G                                                      106

gcaaacctgtgttaaaatcttagctcattcattaattgtgtcatgctgggcaaatcagtc
tctctggcctctttttcctcactcgaaaaatggagacgatgaaaataatgtctcataggt
ttggattaaattaaataatgtaggtacttagtaaatgttctctttcatccctcctttgat
aaatttgccaactgagatttgctgaattacgtctttcttatgccaaaaaaacctaggact
tgttttgatgttaattaaactaaactatatttctgcaagctatcacagaggacagagatt
attttaccgatatactataagtatcatgatttggaaggagtttccctggcgtaggtgccg
```

Fig.7. (3/22)

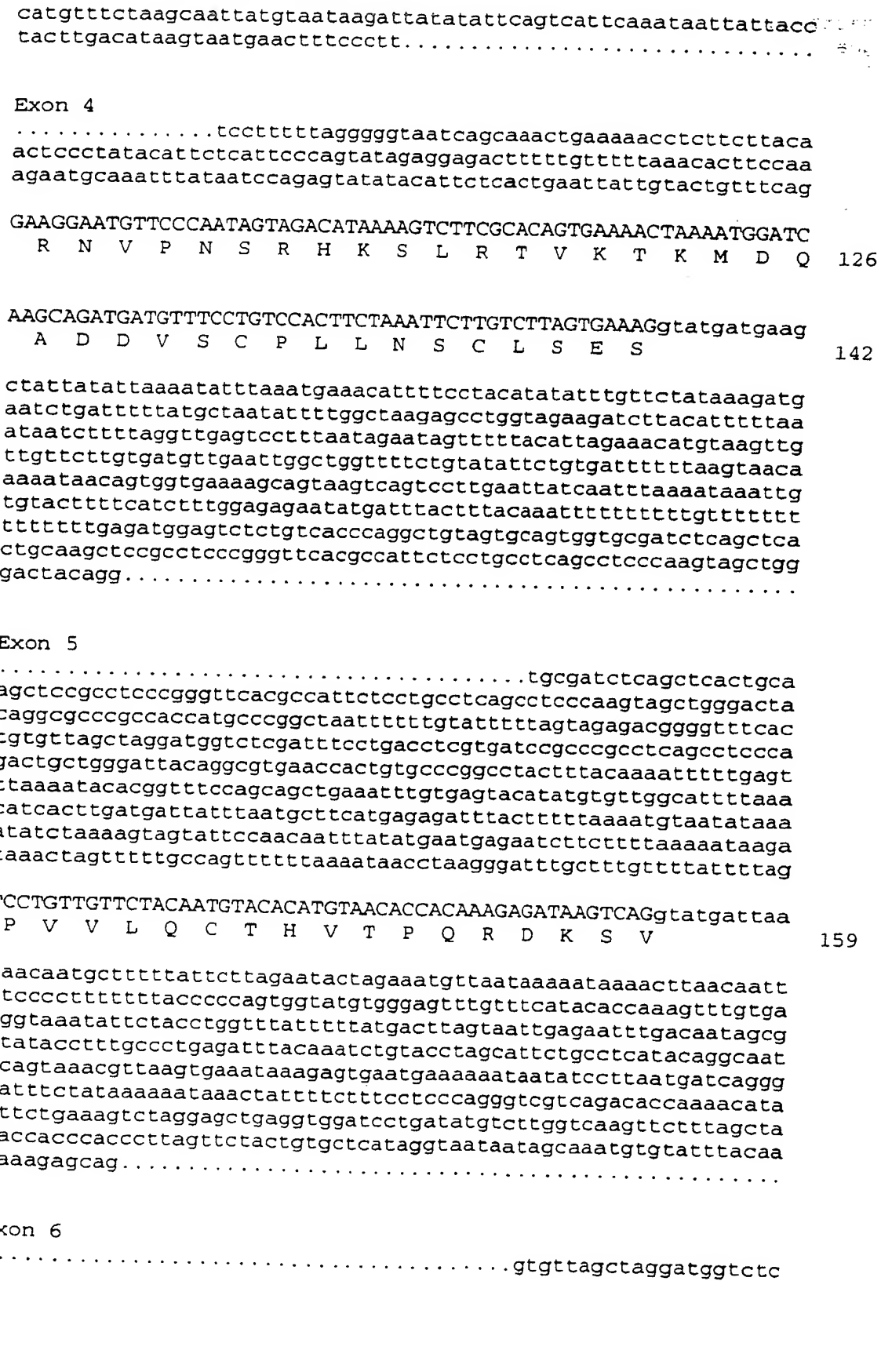

catgtttctaagcaattatgtaataagattatatattcagtcattcaaataattattacc
tacttgacataagtaatgaactttcccttExon 4tcctttttaggggtaatcagcaaactgaaaaacctcttcttaca
actccctatacattctcattcccagtatagaggagacttttgtttttaaacacttccaa
agaatgcaaatttataatccagagtatatacattctcactgaattattgtactgtttcag
GAAGGAATGTTCCCAATAGTAGACATAAAAGTCTTCGCACAGTGAAAACTAAAATGGATC
R N V P N S R H K S L R T V K T K M D Q 126
AAGCAGATGATGTTTCCTGTCCACTTCTAAATTCTTGTCTTAGTGAAAGgtatgatgaag
A D D V S C P L L N S C L S E S 142
ctattatattaaaatatttaaatgaaacattttcctacatatatttgttctataaagatg
aatctgatttttatgctaatattttggctaagagcctggtagaagatcttacatttttaa
ataatcttttaggttgagtcctttaatagaatagtttttacattagaaacatgtaagttg
ttgttcttgtgatgttgaattggctggtttttctgtatattctgtgatttttttaagtaaca
aaaataacagtggtgaaaagcagtaagtcagtccttgaattatcaatttaaaataaattg
tgtacttttcatctttggagagaatatgatttactttacaaatttttttttttgttttttt
ttttttgagatggagtctctgtcacccaggctgtagtgcagtggtgcgatctcagctca
ctgcaagctccgcctcccgggttcacgccattctcctgcctcagcctcccaagtagctgg
gactacaggExon 5tgcgatctcagctcactgca
agctccgcctcccgggttcacgccattctcctgcctcagcctcccaagtagctgggacta
caggcgcccgccaccatgcccggctaatttttttgtatttttagtagagacggggtttcac
tgtgttagctaggatggtctcgatttcctgacctcgtgatccgcccgcctcagcctccca
gactgctgggattacaggcgtgaaccactgtgcccggcctactttacaaaatttttgagt
ttaaaatacacggtttccagcagctgaaatttgtgagtacatatgtgttggcattttaaa
catcacttgatgattatttaatgcttcatgagagatttacttttaaaatgtaatataaa
atatctaaaagtagtattccaacaatttatatgaatgagaatcttcttttaaaaataaga
taaactagtttttgccagtttttttaaaataacctaagggatttgctttgttttatttttag
TCCTGTTGTTCTACAATGTACACATGTAACACCACAAAGAGATAAGTCAGgtatgattaa
P V V L Q C T H V T P Q R D K S V 159
aaacaatgctttttattcttagaatactagaaatgttaataaaaataaaacttaacaatt
ttccccttttttttaccccccagtggtatgtgggagtttgtttcatacaccaaagtttgtga
aggtaaatattctacctggtttatttttatgacttagtaattgagaatttgacaatagcg
ttatacctttgccctgagatttacaaatctgtacctagcattctgcctcatacaggcaat
tcagtaaacgttaagtgaaataaagagtgaatgaaaaaataatatccttaatgatcaggg
catttctataaaaaataaactatttttctttcctcccagggtcgtcagacaccaaaacata
tttctgaaagtctaggagctgaggtggatcctgatatgtcttggtcaagttctttagcta
caccacccacccttagttctactgtgctcataggtaataatagcaaatgtgtatttacaa
gaaagagcagExon 6gtgttagctaggatggtctc

Fig.7. (4/22)

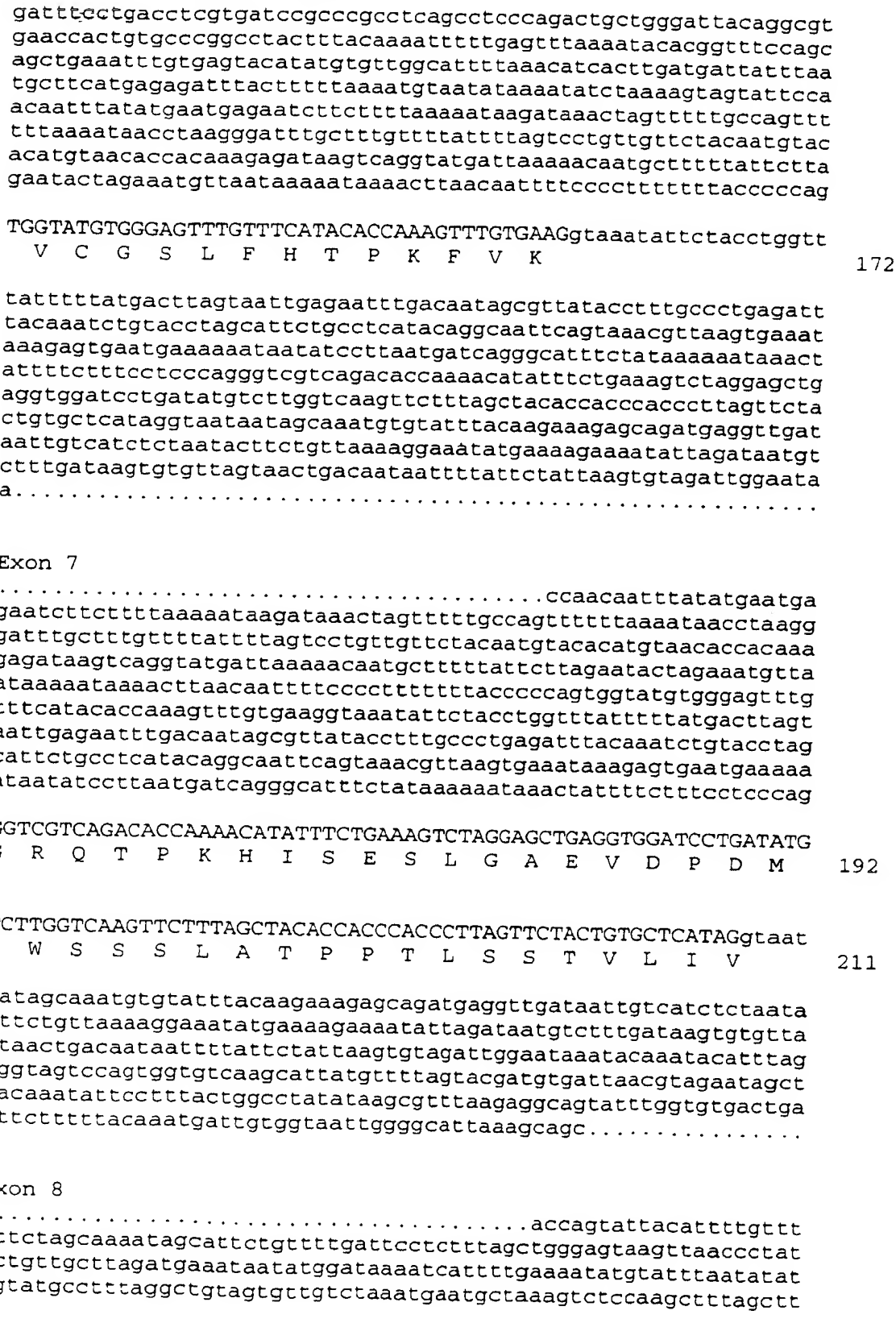

gatttcctgacctcgtgatccgcccgcctcagcctcccagactgctgggattacaggcgt
gaaccactgtgcccggcctactttacaaaatttttgagtttaaaatacacggtttccagc
agctgaaatttgtgagtacatatgtgttggcattttaaacatcacttgatgattatttaa
tgcttcatgagagatttacttttaaaatgtaatataaaatatctaaaagtagtattcca
acaatttatatgaatgagaatcttcttttaaaaataagataaactagtttttgccagttt
tttaaaataacctaagggatttgctttgttttattttagtcctgttgttctacaatgtac
acatgtaacaccacaaagagataagtcaggtatgattaaaaacaatgctttttattctta
gaatactagaaatgttaataaaaataaaacttaacaattttccccttttttttaccccag
TGGTATGTGGGAGTTTGTTTCATACACCAAAGTTTGTGAAGgtaaatattctacctggtt
V C G S L F H T P K F V K
172
tatttttatgacttagtaattgagaatttgacaatagcgttatacctttgccctgagatt
tacaaatctgtacctagcattctgcctcatacaggcaattcagtaaacgttaagtgaaat
aaagagtgaatgaaaaaataatatccttaatgatcagggcatttctataaaaaataaact
attttctttcctcccagggtcgtcagacaccaaaacatatttctgaaagtctaggagctg
aggtggatcctgatatgtcttggtcaagttctttagctacaccacccacccttagttcta
ctgtgctcataggtaataatagcaaatgtgtatttacaagaaagagcagatgaggttgat
aattgtcatctctaatacttctgttaaaaggaaatatgaaaagaaaatattagataatgt
ctttgataagtgtgttagtaactgacaataattttattctattaagtgtagattggaata
a.....
Exon 7
.....ccaacaatttatatgaatga
gaatcttcttttaaaaataagataaactagtttttgccagtttttttaaaataacctaagg
gatttgctttgttttattttagtcctgttgttctacaatgtacacatgtaacaccacaaa
gagataagtcaggtatgattaaaaacaatgctttttattcttagaatactagaaatgtta
ataaaaataaaacttaacaatttttccccttttttttaccccagtggtatgtgggagtttg
tttcatacaccaaagtttgtgaaggtaaatattctacctggtttatttttatgacttagt
aattgagaatttgacaatagcgttatacctttgccctgagatttacaaatctgtacctag
cattctgcctcatacaggcaattcagtaaacgttaagtgaaataaagagtgaatgaaaaa
ataatatccttaatgatcagggcatttctataaaaaataaactatttttctttcctcccag
GGTCGTCAGACACCAAAACATATTTCTGAAAGTCTAGGAGCTGAGGTGGATCCTGATATG
G R Q T P K H I S E S L G A E V D P D M
192
TCTTGGTCAAGTTCTTTAGCTACACCACCCACCCTTAGTTCTACTGTGCTCATAGgtaat
S W S S S L A T P P T L S S T V L I V
211
aatagcaaatgtgtatttacaagaaagagcagatgaggttgataattgtcatctctaata
cttctgttaaaaggaaatatgaaaagaaaatattagataatgtctttgataagtgtgtta
gtaactgacaataattttattctattaagtgtagattggaataaatacaaatacatttag
tggtagtccagtggtgtcaagcattatgttttagtacgatgtgattaacgtagaatagct
tacaaatattcctttactggcctatataagcgtttaagaggcagtatttggtgtgactga
attctttttacaaatgattgtggtaattggggcattaaagcagc.....
Exon 8
.....accagtattacattttgttt
attctagcaaaatagcattctgttttgattcctctttagctgggagtaagttaaccctat
tctgttgcttagatgaaataatatggataaaatcattttgaaaatatgtatttaatatat
agtatgcctttaggctgtagtgttgtctaaatgaatgctaaagtctccaagctttagctt

Fig.7. (5/22)

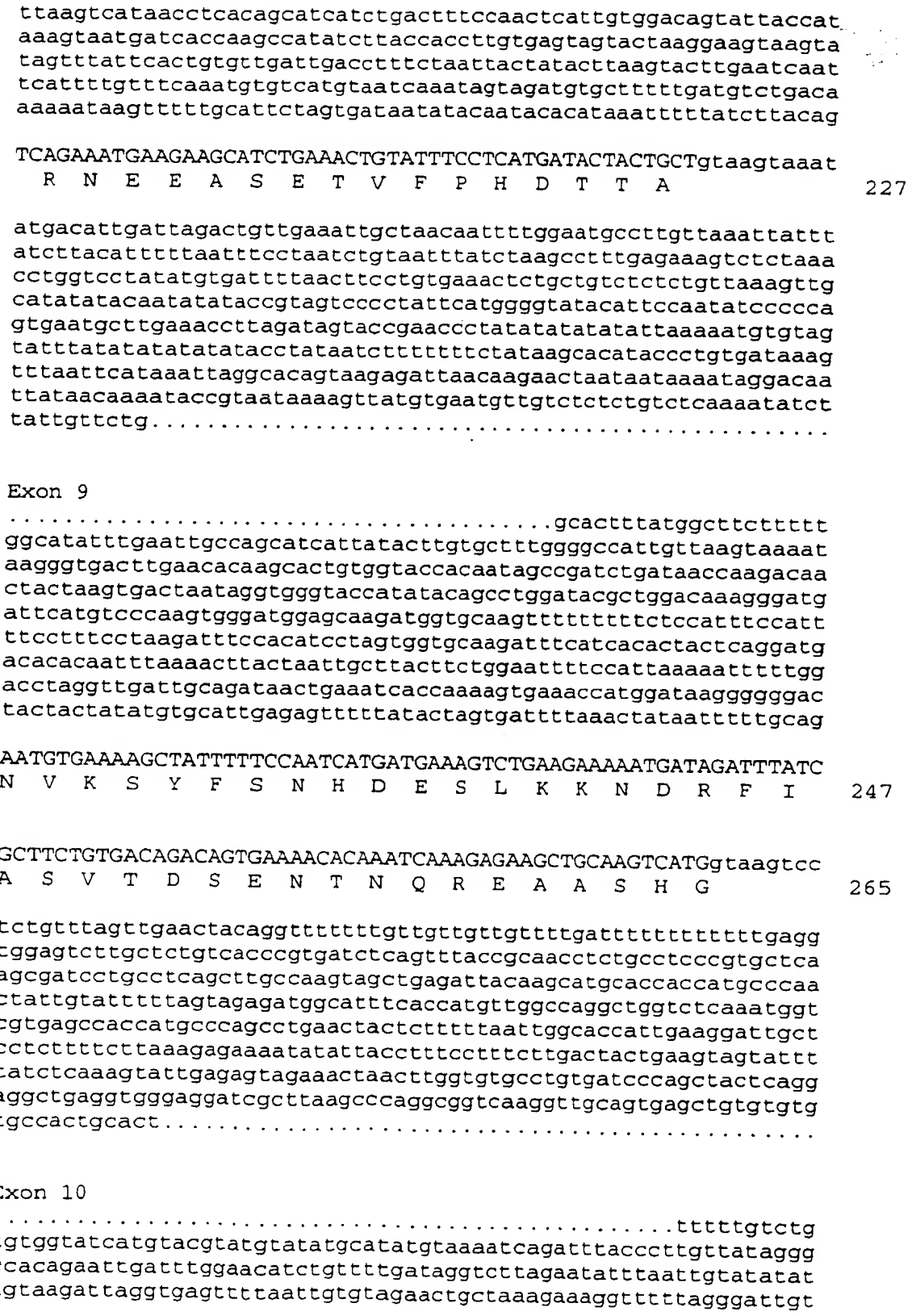

ttaagtcataacctcacagcatcatctgactttccaactcattgtggacagtattaccat
aaagtaatgatcaccaagccatatcttaccaccttgtgagtagtactaaggaagtaagta
tagtttattcactgtgttgattgacctttctaattactatacttaagtacttgaatcaat
tcattttgtttcaaatgtgtcatgtaatcaaatagtagatgtgctttttgatgtctgaca
aaaaataagttttgcattctagtgataatatacaatacacataaatttttatcttacag
TCAGAAATGAAGAAGCATCTGAAACTGTATTTCCTCATGATACTACTGCTgtaagtaaat
R N E E A S E T V F P H D T T A
227
atgacattgattagactgttgaaattgctaacaattttggaatgccttgttaaattattt
atcttacatttttaatttcctaatctgtaatttatctaagcctttgagaaagtctctaaa
cctggtcctatatgtgattttaacttcctgtgaaactctgctgtctctctgttaaagttg
catatatacaatatataccgtagtcccctattcatggggtatacattccaatatccccca
gtgaatgcttgaaaccttagatagtaccgaaccctatatatatatattaaaaatgtgtag
tatttatatatatatatacctataatctttttttctataagcacataccctgtgataaag
tttaattcataaattaggcacagtaagagattaacaagaactaataataaaataggacaa
ttataacaaaataccgtaataaaagttatgtgaatgttgtctctctgtctcaaaatatct
tattgttctg.....
Exon 9
.....gcactttatggcttcttttt
ggcatatttgaattgccagcatcattatacttgtgctttggggccattgttaagtaaaat
aagggtgacttgaacacaagcactgtggtaccacaatagccgatctgataaccaagacaa
ctactaagtgactaataggtgggtaccatatacagcctggatacgctggacaaagggatg
attcatgtcccaagtgggatggagcaagatggtgcaagtttttttttctccatttccatt
ttcctttcctaagatttccacatcctagtggtgcaagatttcatcacactactcaggatg
acacacaatttaaaacttactaattgcttacttctggaattttccattaaaaatttttgg
acctaggttgattgcagataactgaaatcaccaaaagtgaaaccatggataaggggggac
tactactatatgtgcattgagagtttttatactagtgattttaaactataatttttgcag
AATGTGAAAAGCTATTTTTCCAATCATGATGAAAGTCTGAAGAAAAATGATAGATTTATC
N V K S Y F S N H D E S L K K N D R F I
247
GCTTCTGTGACAGACAGTGAAAACACAAATCAAAGAGAAGCTGCAAGTCATGgtaagtcc
A S V T D S E N T N Q R E A A S H G
265
tctgtttagttgaactacaggtttttttgttgttgttgttttgatttttttttttgagg
tggagtcttgctctgtcacccgtgatctcagtttaccgcaacctctgcctcccgtgctca
agcgatcctgcctcagcttgccaagtagctgagattacaagcatgcaccaccatgcccaa
ctattgtatttttagtagagatggcatttcaccatgttggccaggctggtctcaaatggt
cgtgagccaccatgcccagcctgaactactctttttaattggcaccattgaaggattgct
cctcttttcttaaagagaaaatatattacctttccttttcttgactactgaagtagtattt
tatctcaaagtattgagagtagaaactaacttggtgtgcctgtgatcccagctactcagg
aggctgaggtgggaggatcgcttaagcccaggcggtcaaggttgcagtgagctgtgtgtg
tgccactgcact.....
Exon 10
.....tttttgtctg
tgtggtatcatgtacgtatgtatatgcatatgtaaaatcagatttacccttgttataggg
ccacagaattgatttggaacatctgttttgataggtcttagaatatttaattgtatatat
agtaagattaggtgagttttaattgtgtagaactgctaaagaaaggtttttagggattgt

Fig.7. (6/22)

```
tgtatgaataaaaggctttaggttcattggaatcaggggaatcaggctttactagaagaa
caggagaaggggtgactgaccgaaaaataaaatgccaagtactcagaataacccctttaaa
tactgatatgtaatatttagcacattctacataaactgtttctatgagaaaggttgtgag
aataatataaattatatggcttataaaatattaatgtgcttctgttttatactttaacag

GATTTGGAAAAACATCAGGGAATTCATTTAAAGTAAATAGCTGCAAAGACCACATTGGAA
  F  G  K  T  S  G  N  S  F  K  V  N  S  C  K  D  H  I  G  K  285

AGTCAATGCCAAATGTCCTAGAAGATGAAGTATATGAAACAGTTGTAGATACCTCTGAAG
  S  M  P  N  V  L  E  D  E  V  Y  E  T  V  V  D  T  S  E  E  305

AAGATAGTTTTTCATTATGTTTTTCTAAATGTAGAACAAAAAATCTACAAAAAGTAAGAA
  D  S  F  S  L  C  F  S  K  C  R  T  K  N  L  Q  K  V  R  T  325

CTAGCAAGACTAGGAAAAAAATTTTCCATGAAGCAAACGCTGATGAATGTGAAAAATCTA
  S  K  T  R  K  K  I  F  H  E  A  N  A  D  E  C  E  K  S  K  345

AAAACCAAGTGAAAGAAAAATACTCATTTGTATCTGAAGTGGAACCAAATGATACTGATC
  N  Q  V  K  E  K  Y  S  F  V  S  E  V  E  P  N  D  T  D  P  365

CATTAGATTCAAATGTAGCACATCAGAAGCCCTTTGAGAGTGGAAGTGACAAAATCTCCA
  L  D  S  N  V  A  H  Q  K  P  F  E  S  G  S  D  K  I  S  K  385

AGGAAGTTGTACCGTCTTTGGCCTGTGAATGGTCTCAACTAACCCTTTCAGGTCTAAATG
  E  V  V  P  S  L  A  C  E  W  S  Q  L  T  L  S  G  L  N  G  405

GAGCCCAGATGGAGAAAATACCCCTATTGCATATTTCTTCATGTGACCAAAATATTTCAG
  A  Q  M  E  K  I  P  L  L  H  I  S  S  C  D  Q  N  I  S  E  425

AAAAAGACCTATTAGACACAGAGAACAAAAGAAAGAAAGATTTTCTTACTTCAGAGAATT
  K  D  L  L  D  T  E  N  K  R  K  K  D  F  L  T  S  E  N  S  445

CTTTGCCACGTATTTCTAGCCTACCAAAATCAGAGAAGCCATTAAATGAGGAAACAGTGG
  L  P  R  I  S  S  L  P  K  S  E  K  P  L  N  E  E  T  V  V  465

TAAATAAGAGAGATGAAGAGCAGCATCTTGAATCTCATACAGACTGCATTCTTGCAGTAA
  N  K  R  D  E  E  Q  H  L  E  S  H  T  D  C  I  L  A  V  K  485

AGCAGGCAATATCTGGAACTTCTCCAGTGGCTTCTTCATTTCAGGGTATCAAAAAGTCTA
  Q  A  I  S  G  T  S  P  V  A  S  S  F  Q  G  I  K  K  S  I  505

TATTCAGAATAAGAGAATCACCTAAAGAGACTTTCAATGCAAGTTTTTCAGGTCATATGA
  F  R  I  R  E  S  P  K  E  T  F  N  A  S  F  S  G  H  M  T  525

CTGATCCAAACTTTAAAAAAGAAACTGAAGCCTCTGAAAGTGGACTGGAAATACATACTG
  D  P  N  F  K  K  E  T  E  A  S  E  S  G  L  E  I  H  T  V  545

TTTGCTCACAGAAGGAGGACTCCTTATGTCCAAATTTAATTGATAATGGAAGCTGGCCAG
  C  S  Q  K  E  D  S  L  C  P  N  L  I  D  N  G  S  W  P  A  565

CCACCACCACACAGAATTCTGTAGCTTTGAAGAATGCAGGTTTAATATCCACTTTGAAAA
  T  T  T  Q  N  S  V  A  L  K  N  A  G  L  I  S  T  L  K  K  585

AGAAAACAAATAAGTTTATTTATGCTATACATGATGAAACATCTTATAAAGGAAAAAAAA
```

Fig.7. (7/22)

```
  K  T  N  K  F  I  Y  A  I  H  D  E  T  F  Y  K  G  K  K  I  605

TACCGAAAGACCAAAAATCAGAACTAATTAACTGTTCAGCCCAGTTTGAAGCAAATGCTT
  P  K  D  Q  K  S  E  L  I  N  C  S  A  Q  F  E  A  N  A  F  625

TTGAAGCACCACTTACATTTGCAAATGCTGATTCAGgtacctctgtctttttttttttgt
  E  A  P  L  T  F  A  N  A  D  S  G                          637

aaatagtacatatagttttatagatgacgattccttctgtgtttttttctgctttttaaa
atcttcatatcttatatttaatcttaggcatcatctgtatacatgattgtttaggtcttt
aattaccagtgtttagaatcaggtcactcaaacatggtagataagtttgcatagtttgtg
tatatccatcactcttgagacagttttattttaagttccggggtacatgtgcaggatgtg
caggtttgttacataagtaaacgtatgccatgttggtttgctgcacctgtcaacccttca
cctgagtattaagcccagcatgcattagctatttttcctggtgctctccttcccccccaca
caccccacctcctgacagaccctagtgtgtgttgttcccctccctgtgtccgtgtgttc
tcattgttcagctcccacttatgagtgagaacatgt.....

Exon 11

ctgaaaaatagagcatatttaggattctttctgctttaaatttgacattcagttattttc
atgtaatttgtgttttgagcactacctttttaattaatttatttatttttattttttagag
actgtctcattctgttacctagtctggagtgcactagtgtgatctcagctcaccgtagcc
tcaccctcctgggctcaagcagtccttgcacctcaccctcctgagtaactggcaccacag
gcatacaccaccacacccagctaatttttattttttcatagagtcatggtctcactatgtt
gcccaggctagtctcgaactcctgggctcaagcagtcttcctgcctcagcctcccaaaag
tgctgagattacaggcatgagccactgtgcccaaacactacctttttaacttagtgaaaa
atatttagtgaatgtgattgatggtactttaattttgtcactttgtgtttttatgtttag

GTTTATTGCATTCTTCTGTGAAAAGAAGCTGTTCACAGAATGATTCTGAAGAACCAACTT
  L  L  H  S  S  V  K  R  S  C  S  Q  N  D  S  E  E  P  T  L  657

TGTCCTTAACTAGCTCTTTTGGGACAATTCTGAGGAAATGTTCTAGAAATGAAACATGTT
  S  L  T  S  S  F  G  T  I  L  R  K  C  S  R  N  E  T  C  S  677

CTAATAATACAGTAATCTCTCAGGATCTTGATTATAAAGAAGCAAAATGTAATAAGGAAA
  N  N  T  V  I  S  Q  D  L  D  Y  K  E  A  K  C  N  K  E  K  697

AACTACAGTTATTTATTACCCCAGAAGCTGATTCTCTGTCATGCCTGCAGGAAGGACAGT
  L  Q  L  F  I  T  P  E  A  D  S  L  S  C  L  Q  E  G  Q  C  717

GTGRAAATGATCCAAAAAGCAAAAAAGTTTCAGATATAAAAGAAGAGGTCTTGGCTGCAG
  E  N  D  P  K  S  K  K  V  S  D  I  K  E  E  V  L  A  A  A  737

CATGTCACCCAGTACAACATTCAAAAGTGGAATACAGTGATACTGACTTTCAATCCCAGA
  C  H  P  V  Q  H  S  K  V  E  Y  S  D  T  D  F  Q  S  Q  K  757

AAAATCTTTTATATGATCATGAAAATGCCAGCACTCTTATTTTAACTCCTACTTCCAAGG
  S  L  L  Y  D  H  E  N  A  S  T  L  I  L  T  P  T  S  K  D  777

ATGTTCTGTCAAACCTAGTCATGATTTCTAGAGGCAAAGAATCATACAAAATGTCAGACA
  V  L  S  N  L  V  M  I  S  R  G  K  E  S  Y  K  M  S  D  K  797

AGCTCAAAGGTAACAATTATGAATCTGATGTTGAATTAACCAAAAATATTCCCATGGAAA
```

Fig.7. (8/22)

L K G N N Y E S D V E L T K N I P M E K 817

AGAATCAAGATGTATGTGCTTTAAATGAAAATTATAAAAACGTTGAGCTGTTGCCACCTG
N Q D V C A L N E N Y K N V E L L P P E 837

AAAAATACATGAGAGTAGCATCACCTTCAAGAAAGGTACAATTCAACCAAAACACAAATC
K Y M R V A S P S R K V Q F N Q N T N L 857

TAAGAGTAATCCAAAAAAATCAAGAAGAAACTACTTCAATTTCAAAAATAACTGTCAATC
R V I Q K N Q E E T T S I S K I T V N P 877

CAGACTCTGAAGAACTTTTCTCAGACAATGAGAATAATTTTGTCTTCCAAGTAGCTAATG
D S E E L F S D N E N N F V F Q V A N E 897

AAAGGAATAATCTTGCTTTAGGAAATACTAAGGAACTTCATGAAACAGACTTGACTTGTG
R N N L A L G N T K E L H E T D L T C V 917

TAAACGAACCCATTTTCAAGAACTCTACCATGGTTTTATATGGAGACACAGGTGATAAAC
N E P I F K N S T M V L Y G D T G D K Q 937

AAGCAACCCAAGTGTCAATTAAAAAAGATTTGGTTTATGTTCTTGCAGAGGAGAACAAAA
A T Q V S I K K D L V Y V L A E E N K N 957

ATAGTGTAAAGCAGCATATAAAAATGACTCTAGGTCAAGATTTAAAATCGGACATCTCCT
S V K Q H I K M T L G Q D L K S D I S L 977

TGAATATAGATAAAATACCAGAAAAAAATAATGATTACATGAACAAATGGGCAGGACTCT
N I D K I P E K N N D Y M N K W A G L L 997

TAGGTCCAATTTCAAATCACAGTTTTGGAGGTAGCTTCAGAACAGCTTCAAATAAGGAAA
G P I S N H S F G G S F R T A S N K E I 1017

TCAAGCTCTCTGAACATAACATTAAGAAGAGCAAAATGTTCTTCAAAGATATTGAAGAAC
K L S E H N I K K S K M F F K D I E E Q 1037

AATATCCTACTAGTTTAGCTTGTGTTGAAATTGTAAATACCTTGGCATTAGATAATCAAA
Y P T S L A C V E I V N T L A L D N Q K 1057

AGAAACTGAGCAAGCCTCAGTCAATTAATACTGTATCTGCACATTTACAGAGTAGTGTAG
K L S K P Q S I N T V S A H L Q S S V V 1077

TTGTTTCTGATTGTAAAAATAGTCATATAACCCCTCAGATGTTATTTTCCAAGCAGGATT
V S D C K N S H I T P Q M L F S K Q D F 1097

TTAATTCAAACCATAATTTAACACCTAGCCAAAAGGCAGAAATTACAGAACTTTCTACTA
N S N H N L T P S Q K A E I T E L S T I 1117

TATTAGAAGAATCAGGAAGTCAGTTTGAATTTACTCAGTTTAGAAAACCAAGCTACATAT
L E E S G S Q F E F T Q F R K P S Y I L 1137

TGCAGAAGAGTACATTTGAAGTGCCTGAAAACCAGATGACTATCTTAAAGACCACTTCTG
Q K S T F E V P E N Q M T I L K T T S E 1157

AGGAATGCAGAGATGCTGATCTTCATGTCATAATGAATGCCCCATCGATTGGTCAGGTAG

Fig.7. (9/22)

```
  E  C  R  D  A  D  L  H  V  I  M  N  A  P  S  I  G  Q  V  D 1177

ACAGCAGCAAGCAATTTGAAGGTACAGTTGAAATTAAACGGAAGTTTGCTGGCCTGTTGA
  S  S  K  Q  F  E  G  T  V  E  I  K  R  K  F  A  G  L  L  K 1197

AAAATGACTGTAACAAAAGTGCTTCTGGTTATTTAACAGATGAAAATGAAGTGGGGTTTA
  N  D  C  N  K  S  A  S  G  Y  L  T  D  E  N  E  V  G  F  R 1217

GGGGCTTTTATTCTGCTCATGGCACAAAACTGAATGTTTCTACTGAAGCTCTGCAAAAAG
  G  F  Y  S  A  H  G  T  K  L  N  V  S  T  E  A  L  Q  K  A 1237

CTGTGAAACTGTTTAGTGATATTGAGAATATTAGTGAGGAAACTTCTGCAGAGGTACATC
  V  K  L  F  S  D  I  E  N  I  S  E  E  T  S  A  E  V  H  P 1257

CAATAAGTTTATCTTCAAGTAAATGTCATGATTCTGTCGTTTCAATGTTTAAGATAGAAA
  I  S  L  S  S  S  K  C  H  D  S  V  V  S  M  F  K  I  E  N 1277

ATCATAATGATAAAACTGTAAGTGAAAAAAATAATAAATGCCAACTGATATTACAAAATA
  H  N  D  K  T  V  S  E  K  N  N  K  C  Q  L  I  L  Q  N  N 1297

ATATTGAAATGACTACTGGCACTTTTGTTGAAGAAATTACTGAAAATTACAAGAGAAATA
  I  E  M  T  T  G  T  F  V  E  E  I  T  E  N  Y  K  R  N  T 1317

CTGAAAATGAAGATAACAAATATACTGCTGCCAGTAGAAATTCTCATAACTTAGAATTTG
  E  N  E  D  N  K  Y  T  A  A  S  R  N  S  H  N  L  E  F  D 1337

ATGGCAGTGATTCAAGTAAAAATGATACTGTTTGTATTCATAAAGATGAAACGGACTTGC
  G  S  D  S  S  K  N  D  T  V  C  I  H  K  D  E  T  D  L  L 1357

TATTTACTGATCAGCACAACATATGTCTTAAATTATCTGGCCAGTTTATGAAGGAGGGAA
  F  T  D  Q  H  N  I  C  L  K  L  S  G  Q  F  M  K  E  G  N 1377

ACACTCAGATTAAAGAAGATTTGTCAGATTTAACTTTTTTGGAAGTTGCGAAAGCTCAAG
  T  Q  I  K  E  D  L  S  D  L  T  F  L  E  V  A  K  A  Q  E 1397

AAGCATGTCATGGTAATACTTCAAATAAAGAACAGTTAACTGCTACTAAAACGGAGCAAA
  A  C  H  G  N  T  S  N  K  E  Q  L  T  A  T  K  T  E  Q  N 1417

ATATAAAAGATTTTGAGACTTCTGATACATTTTTTCAGACTGCAAGTGGGAAAAATATTA
  I  K  D  F  E  T  S  D  T  F  F  Q  T  A  S  G  K  N  I  S 1437

GTGTCGCCAAAGAGTCATTTAATAAAATTGTAAATTTCTTTGATCAGAAACCAGAAGAAT
  V  A  K  E  L  F  N  K  I  V  N  F  F  D  Q  K  P  E  E  L 1457

TGCATAACTTTTCCTTAAATTCTGAATTACATTCTGACATAAGAAAGAACAAAATGGACA
  H  N  F  S  L  N  S  E  L  H  S  D  I  R  K  N  K  M  D  I 1477

TTCTAAGTTATGAGGAAACAGACATAGTTAAACACAAAATACTGAAAGAAAGTGTCCCAG
  L  S  Y  E  E  T  D  I  V  K  H  K  I  L  K  E  S  V  P  V 1497

TTGGTACTGGAAATCAACTAGTGACCTTCCAGGGACAACCCGAACGTGATGAAAAGATCA
  G  T  G  N  Q  L  V  T  F  Q  G  Q  P  E  R  D  E  K  I  K 1517

AAGAACCTACTCTGTTGGGTTTTCATACAGCTAGCGGGAAAAAAGTTAAAATTGCAAAGG
```

Fig.7. (10/22)


```
  E  P  T  L  L  G  F  H  T  A  S  G  K  K  V  K  I  A  K  E 1537

AATCTTTGGACAAAGTGAAAAACCTTTTTGATGAAAAAGAGCAAGGTACTAGTGAAATCA
  S  L  D  K  V  K  N  L  F  D  E  K  E  Q  G  T  S  E  I  T 1557

CCAGTTTTAGCCATCAATGGGCAAAGACCCTAAAGTACAGAGAGGCCTGTAAAGACCTTG
  S  F  S  H  Q  W  A  K  T  L  K  Y  R  E  A  C  K  D  L  E 1577

AATTAGCATGTGAGACCATTGAGATCACAGCTGCCCCAAAGTGTAAAGAAATGCAGAATT
  L  A  C  E  T  I  E  I  T  A  A  P  K  C  K  E  M  Q  N  S 1597

CTCTCAATAATGATAAAAACCTTGTTTCTATTGAGACTGTGGTGCCACCTAAGCTCTTAA
  L  N  N  D  K  N  L  V  S  I  E  T  V  V  P  P  K  L  L  S 1617

GTGATAATTTATGTAGACAAACTGAAAATCTCAAAACATCAAAAAGTATCTTTTTGAAAG
  D  N  L  C  R  Q  T  E  N  L  K  T  S  K  S  I  F  L  K  V 1637

TTAAAGTACATGAAAATGTAGAAAAAGAAACAGCAAAAAGTCCTGCAACTTGTTACACAA
  K  V  H  E  N  V  E  K  E  T  A  K  S  P  A  T  C  Y  T  N 1657

ATCAGTCCCCTTATTCAGTCATTGAAAATTCAGCCTTAGCTTTTTACACAAGTTGTAGTA
  Q  S  P  Y  S  V  I  E  N  S  A  L  A  F  Y  T  S  C  S  R 1677

GAAAAACTTCTGTGAGTCAGACTTCATTACTTGAAGCAAAAAAATGGCTTAGAGAAGGAA
  K  T  S  V  S  Q  T  S  L  L  E  A  K  K  W  L  R  E  G  I 1697

TATTTGATGGTCAACCAGAAAGAATAAATACTGCAGATTATGTAGGAAATTATTTGTATG
  F  D  G  Q  P  E  R  I  N  T  A  D  Y  V  G  N  Y  L  Y  E 1717

AAAATAATTCAAACAGTACTATAGCTGAAAATGACAAAAATCATCTCTCCGAAAAACAAG
  N  N  S  N  S  T  I  A  E  N  D  K  N  H  L  S  E  K  Q  D 1737

ATACTTATTTAAGTAACAGTAGCATGTCTAACAGCTATTCCTACCATTCTGATGAGGTAT
  T  Y  L  S  N  S  S  M  S  N  S  Y  S  Y  H  S  D  E  V  Y 1757

ATAATGATTCCAGGATACTCTCAAAAAATAACTTGGATTCTGGTATTGAGCCAGTATTGA
  N  D  S  R  I  L  S  K  N  K  L  D  S  G  I  E  P  V  L  K 1777

AGAATGTTGAAGATCAAAAAAACACTAGTTTTTCCAAAGTAATATCCAATGTAAAAGATG
  N  V  E  D  Q  K  N  T  S  F  S  K  V  I  S  N  V  K  D  A 1797

CAAATGCATACCCACAAACTGTAAATGAAGATATTTGCGTTGAGGAACTTGTGACTAGCT
  N  A  Y  P  Q  T  V  N  E  D  I  C  V  E  E  L  V  T  S  S 1817

CTTCACCCTGCAAAAATAAAAATGCAGCCATTAAATTGTCCATATCTAATAGTAATAATT
  S  P  C  K  N  K  N  A  A  I  K  L  S  I  S  N  S  N  N  F 1837

TTGAGGTAGGGCCACCTGCATTTAGGATAGCCAGTGGTAAAATCGTTTGTGTTTCACATG
  E  V  G  P  P  A  F  R  I  A  S  G  K  I  R  L  C  S  H  E 1857

AAACAATTAAAAAAGTGAAAGACATATTTACAGACAGTTTCAGTAAAGTAATTAAGGAAA
  T  I  K  K  V  K  D  I  D  S  F  S  K  V  I  F  T  K  E  N 1877

ACAACGAGAATAAATCAAAAATTTGCCAAACGAAAATTATGGCAGGTTGTTACGAGGCAT
```

Fig.7. (11/22)

```
                                                                N  E  N  K  S  K  I  C  Q  T  K  I  M  A  G  C  Y  E  A  L 1897

TGGATGATTCAGAGGATATTCTTCATAACTCTCTAGATAATGATGAATGTAGCACGCATT
  D  D  S  E  D  I  L  H  N  S  L  D  N  D  E  C  S  M  H  S 1917

CACATAAGGTTTTTGCTGACATTCAGAGTGAAGAAATTTTACAACATAACCAAAATATGT
  H  K  V  F  A  D  I  Q  S  E  E  I  L  Q  H  N  Q  N  M  S 1937

CTGGATTGGAGAAAGTTTCTAAAATATCACCTTGTGATGTTAGTTTGGAAACTTCAGATA
  G  L  E  K  V  S  K  I  S  P  C  D  V  S  L  E  T  S  D  I 1957

TATGTAAATGTAGTATAGGGAAGCTTCATAAGTCAGTCTCATCTGCAAATACTTGTGGGA
  C  K  C  S  I  G  K  L  H  K  S  V  S  S  A  N  T  C  G  I 1977

TTTTTAGCACAGCAAGTGGAAAATCTGTCCAGGTATCAGATGCTTCATTACAAAACGCAA
  F  S  T  A  S  G  K  S  V  Q  V  S  D  A  S  L  Q  N  A  R 1997

GACAAGTGTTTTCTGAAATAGAAGATAGTACCAAGCAAGTCTTTTCCAAAGTATTGTTTA
  Q  V  F  S  E  I  E  D  S  T  K  Q  V  F  S  K  V  L  F  K 2017

AAAGTAACGAACATTCAGACCAGCTCACAAGAGAAGAAAATACTGCTATACGTACTCCAG
  S  N  E  H  S  D  Q  L  T  R  E  E  N  T  A  I  R  T  P  E 2037

AACATTTAATATCCCAAAAAGGCTTTTCATATAATGTGGTAAATTCATCTGCTTTCTCTG
  H  L  I  S  Q  K  G  F  S  Y  N  V  V  N  S  S  A  F  S  G 2057

GATTTAGTACAGCAAGTGGAAAGCAAGTTTCCATTTTAGAAAGTTCCTTACACAAAGTTA
  F  S  T  A  S  G  K  Q  V  S  I  L  E  S  S  L  H  K  V  K 2077

AGGGAGTGTTAGAGGAATTTGATTTAATCAGAACTGAGCATAGTCTTCACTATTCACCTA
  G  V  L  E  E  F  D  L  I  R  T  E  H  S  L  H  Y  S  P  T 2097

CGTCTAGACAAAATGTATCAAAAATACTTCCTCGTGTTGATAAGAGAAACCCAGAGCACT
  S  R  Q  N  V  S  K  I  L  P  R  V  D  K  R  N  P  E  H  C 2117

GTGTAAACTCAGAAATGGAAAAAACCTGCAGTAAAGAATTTAAATTATCAAATAACTTAA
  V  N  S  E  M  E  K  T  C  S  K  E  F  K  L  S  N  N  L  N 2137

ATGTTGAAGGTGGTTCTTCAGAAAATAATCACTCTATTAAAGTTTCTCCATATCTCTCTC
  V  E  G  G  S  S  E  N  N  H  S  I  K  V  S  P  Y  L  S  Q 2157

AATTTCAACAAGACAAACAACAGTTGGTATTAGGAACCAAAGTCTCACTTGTTGAGAACA
  F  Q  Q  D  K  Q  Q  L  V  L  G  T  K  V  S  L  V  E  N  I 2177

TTCATGTTTTGGGAAAAGAACAGGCTTCACCTAAAAACGTAAAAATGGAAATTGGTAAAA
  H  V  L  G  K  E  Q  A  S  P  K  N  V  K  M  E  I  G  K  T 2197

CTGAAACTTTTTCTGATGTTCCTGTGAAAACAAATATAGAAGTTTGTTCTACTTACTCCA
  E  T  F  S  D  V  P  V  K  T  N  I  E  V  C  S  T  Y  S  K 2217

AAGATTCAGAAAACTACTTTGAAACAGAAGCAGTAGAAATTGCTAAAGCTTTTATGGAAG
  D  S  E  N  Y  F  E  T  E  A  V  E  I  A  K  A  F  M  E  D 2237

ATGATGAACTGACAGATTCTAAACTGCCAAGTCATGCCACACATTCTCTTTTTACATGTC
```

Fig.7. (12/22)

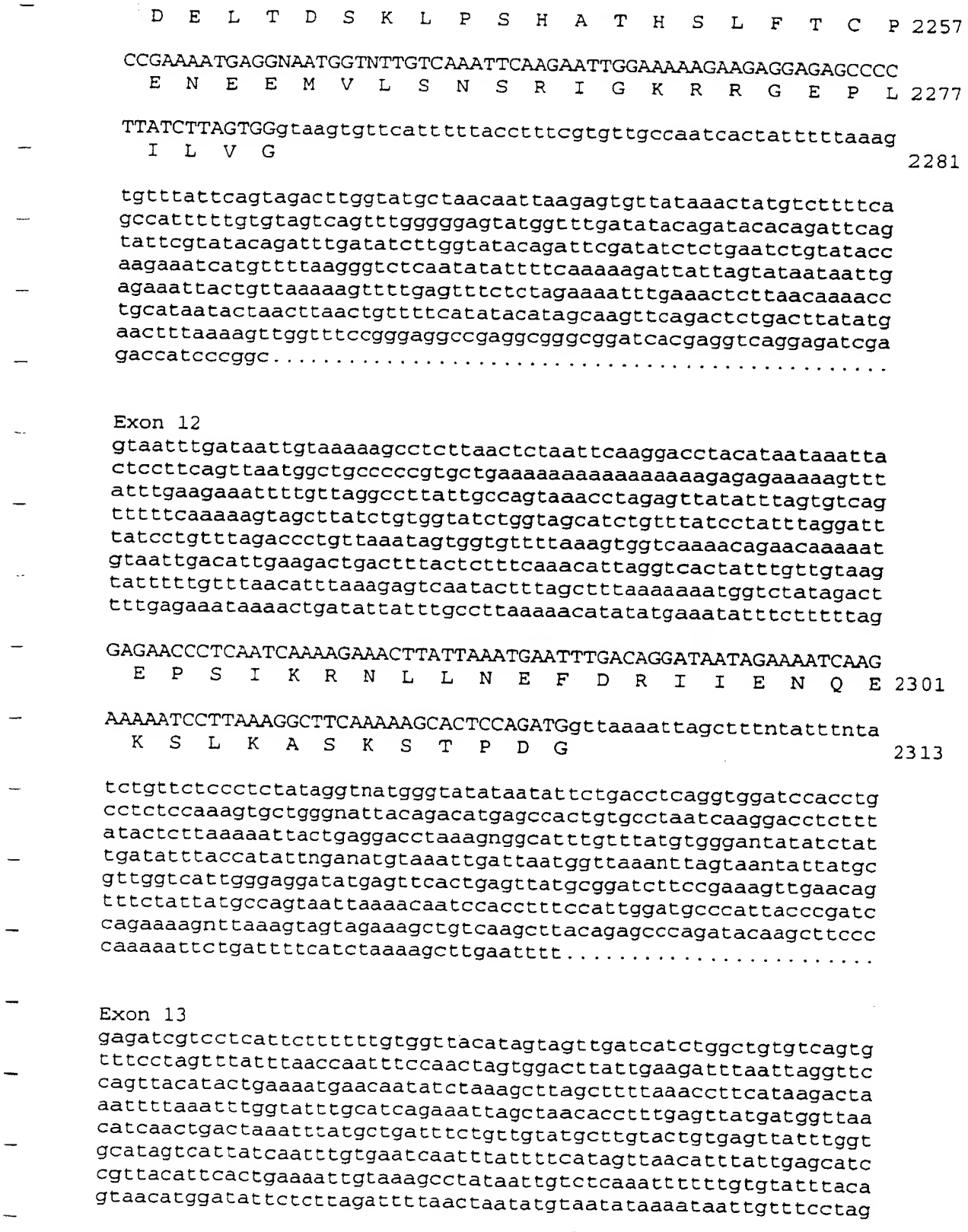

D E L T D S K L P S H A T H S L F T C P 2257
CCGAAAATGAGGNAATGGTNTTGTCAAATTCAAGAATTGGAAAAAGAAGAGGAGAGCCCC
E N E E M V L S N S R I G K R R G E P L 2277
TTATCTTAGTGGGtaagtgttcatttttacctttcgtgttgccaatcactatttttaaag
I L V G
2281
tgtttattcagtagacttggtatgctaacaattaagagtgttataaactatgtcttttca
gccatttttgtgtagtcagtttgggggagtatggtttgatatacagatacacagattcag
tattcgtatacagatttgatatcttggtatacagattcgatatctctgaatctgtatacc
aagaaatcatgttttaagggtctcaatatattttcaaaaagattattagtataataattg
agaaattactgttaaaaagttttgagtttctctagaaaatttgaaactcttaacaaaacc
tgcataatactaacttaactgttttcatatacatagcaagttcagactctgacttatatg
aactttaaaagttggtttccgggaggccgaggcgggcggatcacgaggtcaggagatcga
gaccatcccggc.....
Exon 12
gtaatttgataattgtaaaaagcctcttaactctaattcaaggacctacataataaatta
ctccttcagttaatggctgcccccgtgctgaaaaaaaaaaaaaaaaagagagaaaaagttt
atttgaagaaatttttgttaggccttattgccagtaaacctagagttatatttagtgtcag
tttttcaaaaagtagcttatctgtggtatctggtagcatctgtttatcctatttaggatt
tatcctgtttagaccctgttaaatagtggtgttttaaagtggtcaaaacagaacaaaaat
gtaattgacattgaagactgactttactctttcaaacattaggtcactatttgttgtaag
tatttttgtttaacatttaaagagtcaatactttagctttaaaaaaatggtctatagact
tttgagaaataaaactgatattatttgccttaaaaacatatatgaaatatttctttttag
GAGAACCCTCAATCAAAAGAAACTTATTAAATGAATTTGACAGGATAATAGAAAATCAAG
E P S I K R N L L N E F D R I I E N Q E 2301
AAAAATCCTTAAAGGCTTCAAAAAGCACTCCAGATGgttaaaattagctttntatttnta
K S L K A S K S T P D G
2313
tctgttctccctctataggtnatgggtatataatattctgacctcaggtggatccacctg
cctctccaaagtgctgggnattacagacatgagccactgtgcctaatcaaggacctcttt
atactcttaaaaattactgaggacctaaagnggcatttgtttatgtgggantatatctat
tgatatttaccatattnganatgtaaattgattaatggttaaanttagtaantattatgc
gttggtcattgggaggatatgagttcactgagttatgcggatcttccgaaagttgaacag
tttctattatgccagtaattaaaacaatccacctttccattggatgcccattacccgatc
cagaaaagnttaaagtagtagaaagctgtcaagcttacagagcccagatacaagcttccc
caaaaattctgattttcatctaaaagcttgaatttt.....
Exon 13
gagatcgtcctcattctttttgtggttacatagtagttgatcatctggctgtgtcagtg
tttcctagtttatttaaccaatttccaactagtggacttattgaagatttaattaggttc
cagttacatactgaaaatgaacaatatctaaagcttagcttttaaaccttcataagacta
aattttaaatttggtatttgcatcagaaattagctaacacctttgagttatgatggttaa
catcaactgactaaatttatgctgatttctgttgtatgcttgtactgtgagttatttggt
gcatagtcattatcaatttgtgaatcaatttattttcatagttaacatttattgagcatc
cgttacattcactgaaaattgtaaagcctataattgtctcaaatttttgtgtatttaca
gtaacatggatattctcttagattttaactaatatgtaatataaaataattgtttcctag

Fig.7. (13/22)

```
GCACAATAAAAGATCGAAGATTGTTTATGCATCATGTTTCTTTAGAGCCGATTACCTGTG
  T  I  K  D  R  R  L  F  M  H  H  V  S  L  E  P  I  T  C  V 2333

TACCCTTTCGgtaagacatgtttaaatttttctaaattctaatacagtatgagaaaagtc
  P  F  R                                                    2336

tcgtttttataaatgaacatttctaaaaataatgacactaacgttaagaagttaacactt
cccgtttttataaaatttataaaatactttggtagtattttatagtgctgttcatatcatt
attttatttttttaatttttatgacagctttgtaaagtagacagattttattctaattttat
ggatgaagtactaaggttgagaggaattaaggaaattgctccgaatcagttaacaaaaag
attgcagatattaaaaatatccttttatctctcctctctaaacctttaaaaaagtactaa
gatagttttttttaatgtataattcccaaggacaatgatgagaagaaacaacaaaagtttg
gaagccaaaaacataaaggatttagtaagcatgagaaagctaaaacctgacactagagca
aacagagatg.....

Exon 14
ttcaaaggtaggcaagatttttgggctaaataaaaagggcactttaaaaaaggtataaat
aggtagaagagagaaaagggagcgaggtgggataattgaaagaggggatctcctgtggag
actgaggtattaggcggagtagagagttcaggtgaagatgtgaaggtgagagaagaggat
gggtagacatttccctggtgaaggaggtaaggagtactatgatggaattagaggggacac
actgagagggtccacacttgacagactctcttctattatgtgttatgtgaggtagattgt
aaagtcaaaggctagccttgaaaaatgtgatattgttttggaatggcaaccatggtgaat
acaaaacagttaccagaatagtatcaccatgtagcaaatgagggtctgcaacaaaggcat
attcctaaatatttatatgtgtactagtcaataaacttatatattttctccccattgcag

CACAACTAAGGAACGTCAAGAGATACAGAATCCAAATTTTACCGCACCTGGTCAAGAATT
 T  T  K  E  R  Q  E  I  Q  N  P  N  F  T  A  P  G  Q  E  F 2356

TCTGTCTAAATCTCATTTGTATGAACATCTGACTTTGGAAAAATCTTCAAGCAATTTAGC
 L  S  K  S  H  L  Y  E  H  L  T  L  E  K  S  S  S  N  L  A 2376

AGTTTCAGGACATCCATTTTATCAAGTTTCTGCTACAAGAAATGAAAAAATGAGACACTT
 V  S  G  H  P  F  Y  Q  V  S  A  T  R  N  E  K  M  R  H  L 2396

GATTACTACAGGCAGACCAACCAAAGTCTTTGTTCCACCTTTTAAAACTAAATCACATTT
 I  T  T  G  R  P  T  K  V  F  V  P  P  F  K  T  K  S  H  F 2416

TCACAGAGTTGAACAGTGTGTTAGGAATATTAACTTGGAGGAAAACAGACAAAAGCAAAA
 H  R  V  E  Q  C  V  R  N  I  N  L  E  E  N  R  Q  K  Q  N 2436

CATTGATGGACATGGCTCTGATGATAGTAAAAATAAGATTAATGACAATGAGATTCATCA
 I  D  G  H  G  S  D  D  S  K  N  K  I  N  D  N  E  I  H  Q 2456

GTTTAACAAAAACAACTCCAATCAAGCAGCAGCTGTAACTTTCACAAAGTGTGAAGAAGA
 F  N  K  N  N  S  N  Q  A  A  A  V  T  F  T  K  C  E  E  E 2476

ACCTTTAGgtattgtatgacaatttgtgtgatgaatttttgcctttcagttagatatttc
 P  L  D                                                    2479

cgttgttaaataatgtcctgatggttttccccctttggtggtggtaattttaaagccctt
tttaatgttttagattttctaaatccaaagattaggtttaaattattctaatgtttcttt
caaagataacttcttgtggacttgttaaaaaaaattagacacacaatct.....
```

Fig.7. (14/22)

Exon15

```
.....................................................aacccactgtgcctgg
ccaggggttgtgctttttaaatttcaattttatttttgctaagtatttattctttgatag

ATTTAATTACAAGTCTTCAGAATGCCAGAGATATACAGGATATGCGAATTAAGAAGAAAC
  L  I  T  S  L  Q  N  A  R  D  I  Q  D  M  R  I  K  K  K  Q 2499

AAAGGCAACGCGTCTTTCCACAGCCAGGCAGTCTGTATCTTGCAAAAACATCCACTCTGC
  R  Q  R  V  F  P  Q  P  G  S  L  Y  L  A  K  T  S  T  L  P 2519

CTCGAATCTCTCTGAAAGCAGCAGTAGGAGGCCAAGTTCCCTCTGCgtgtccccataaac
  R  I  S  L  K  A  A  V  G  G  Q  V  P  S  A                2534

aggtatgtgtttgnctacaatactgatggcttttatgacagagtgtaattttatttcatt
aactagtatcctacaaatggctttgtttaaagaatgaacacattagtgcaggaatggatg
aatgaaatcatccatattttcctaattaagccctgcagtggcagcctctggccccttgct
aggcctgccctcatccctactaaagtgatctgtgccttcccaaattactacttctttttcc
ccccttcaaatctttcttattttgtcattgtaaatgctctcagctaggtgttaaagtagt
cttactgatattcaaatgtgaataactgatagccctgaaccttctatgagctatttatat
tttccaaagaggattctccttaagccaatattatctaggtagaattttaggcaatggaga
ggtgaaaataatattgatgacattaatagctaactttgagcatttt.............
```

Exon 16

```
gtctgactccagagtcaaactctgaacaaacaaaaagacactttgggttagatatcctgg
ggtgaaagcaagcactttgaaagtaagccaagcctgtgtacagatctgaccacctgaggt
cacattccctaaaatacttaaacttctccctttttgtttcccatctaagtttttgaactta
agagattttgtaaaacatcacatttttttatcctcacagtaccttcctatggcagattta
gcaggaggcgtataaacggggtggaaaaggtacagcagactgtggaatgtatggatcatt
tatattacattaaaatttttagtttctagtaaataacttaaatgtttttgtagtgaagat
tctagtagttaatgaaaatttttggtaaattcagttttggtttgttataattgtttttat
tgtgtgatacatgtttactttaaattgtttttctttttttgtgtgtgtttatttttgtgtag

GTGTTCTCATAAACAGCTGTATACGTATGGCGTTTCTAAACATTGCATAAAAATTAACAG
  C  S  H  K  Q  L  Y  T  Y  G  V  S  K  H  C  I  K  I  N  S 2554

CAAAAATGCAGAGTCTTTTCAGTTTCACACTGAAGATTATTTTGGTAAGGAAAGTTTATG
  K  N  A  E  S  F  Q  F  H  T  E  D  Y  F  G  K  E  S  L  W 2574

GACTGGAAAAGGAATACAGTTGGCTGATGGTGGATGGCTCATACCCTCCAATGATGGAAA
  T  G  K  G  I  Q  L  A  D  G  G  W  L  I  P  S  N  D  G  K 2594

GGCTGGAAAAGAAGAATTTTATAGgtactctatgcaaaaagattgtgtgttaacttttat
  A  G  K  E  E  F  Y  R                                      2602

gtattccctcatccctctttcttctcttaactgtctctcgaactaaaaagttggctagaa
atcaaatttttatgcatttaattgttttaagtgcattatggttaagcattctgtagaagt
cttttgaaaagtgctgtttgtcctggggtttaatgaactggatttttcttgatttgggaca
ttttttcttaggcatttataaatatagcccaatttataaagttaaatttggccgggtacag
tggctcatgcctgtaatcccagcactttgggaggccgaggcgggtagatcacctgaggtc
aggagttcgagaccagcctggccaacgtggcgaaaccccatctctactaaaagtacaaga
```

Fig.7. (15/22)

```
actatctgggcgtggtggcaggcacctgtaatcccggctactctggaggctgaggcagga
gaatcgcttgaacctgggaggcag.....

Exon 17
.....ttgttttgtttttatatttt
gagatagggtctcactcttgtccaggctggagtgcagtggcactatcatggctcactgca
gcctcaacctcctgggctcaagcaatcctcccaccacagcctcctaagtagctgggacca
cagatgtgagctaccactcttggctgatttttttattatttttgtagagatgtggggg
tctcactatgttgcctaggctggtctcaaacttctggcctcaagcaatcctcctgcctca
gcttcccaaaatgctgggagtataggcatgagccaccatgctcagcaatgaagtttttat
cagtatgatactttgatacatgtcaaataatttttctgaaattatattgtagatcatatga
actcataaaaacttaatgatcttgaacaatgtagtttttgtacagagaatagttgtagtt
gttgaattcagtatcatcctatgtggtttttatgataatattctacttttatttgttcag

GGCTCTGTGTGACACTCCAGGTGTGGATCCAAAGCTTATTTCTAGAATTTGGGTTTATAA
 A  L  C  D  T  P  G  V  D  P  K  L  I  S  R  I  W  V  Y  N  2622

TCACTATAGATGGATCATATGGAAACTGGCAGCTATGGAATGTGCCTTTCCTAAGGAATT
 H  Y  R  W  I  I  W  K  L  A  A  M  E  C  A  F  P  K  E  F  2642

TGCTAATAGATGCCTAAGCCCAGAAAGGGTGCTTCTTCAACTAAAATACAGgcaagttta
 A  N  R  C  L  S  P  E  R  V  L  L  Q  L  K  Y  R           2659

aagcattacattacgtaatcatatacggcagtatgggttaaggtttctgtgtagtctgtg
acttccatgtcaaaatgttgcacaagccagttgtcagtgacagttgccatcccacactgc
tgttctcctgtcatccctagccccatttaagagagatcacacattcatgcattgcttgc
ttccctctttccccacccccctccttaacctcttgatgtatgagaagaatatgagttacta
atttgatccactatttggggattgctaataaagcatttttgcattttatttttgctttt
taaaaataattgatattttaacaatatgaaacaatatattcctagctacaaaatttttaa
ttctcagtatttcttagataaattcagttttattctcagttattcagtgacttgtttaa
acagtggaattctagagtcacacttcctaaaatatgcatttttgttttcacttttagata
tgatacggaaa.....

Exon 18
.....caactaaaatacaggcaagt
ttaaagcattacattacgtaatcatatacggcagtatgggttaaggtttctgtgtagtct
gtgacttccatgtcaaaatgttgcacaagccagttgtcagtgacagttgccatcccacac
tgctgttctcctgtcatccctagccccatttaagagagatcacacattcatgcattgct
tgcttccctctttccccacccccctccttaacctcttgatgtatgagaagaatatgagtta
ctaatttgatccactatttggggattgctaataaagcatttttgcatttttatttttgct
ttttaaaaataattgatattttaacaatatgaaacaatatattcctagctacaaaatttt
taattctcagtatttcttagataaattcagttttattctcagttattcagtgacttgtt
taaacagtggaattctagagtcacacttcctaaaatatgcatttttgttttcacttttag

ATATGATACGGAAATTGATAGAAGCAGAAGATCGGCTATAAAAAAGATAATGGAAAGGGA
 Y  D  T  E  I  D  R  S  R  R  S  A  I  K  K  I  M  E  R  D  2679

TGACACAGCTGCAAAAACACTTGTTCTCTGTGTTTCTGACATAATTTCATTGAGCGCAAA
 D  T  A  A  K  T  L  V  L  C  V  S  D  I  I  S  L  S  A  N  2699

TATATCTGAAACTTCTAGCAATAAAACTAGTAGTGCAGATACCCAAAAAGTGGCCATTAT
 I  S  E  T  S  S  N  K  T  S  S  A  D  T  Q  K  V  A  I  I  2719
```

Fig.7. (16/22)

```
TGAACTTACAGATGGGTGGTATGCTGTTAAGGCCCAGTTAGATCCTCCCCTCTTAGCTGT
 E  L  T  D  G  W  Y  A  V  K  A  Q  L  D  P  P  L  L  A  V  2739

CTTAAAGAATGGCAGACTGACAGTTGGTCAGAAGATTATTCTTCATGGAGCAGAACTGGT
 L  K  N  G  R  L  T  V  G  Q  K  I  I  L  H  G  A  E  L  V  2759

GGGCTCTCCTGATGCCTGTACACCTCTTGAAGCCCCAGAATCTCTTATGTTAAAGgtaaa
 G  S  P  D  A  C  T  P  L  E  A  P  E  S  L  M  L  K        2777

ttaatttgcactcttggtaaaaatcagtcattgattcagttaaattctagaagttttaca
tttaaattttaaatgcttactaaggatgctcaatttcttagatgtactgataattttagt
ataaaaagcatattcttcagacagttaaagtttttgtgcagtttttgggaggtccagaga
tctttcttgagcttaaataatgcatttccaattaaaaagcaaaataaatttgcaccattt
gattttggtatctgtagcttgctgccctcttgttctcatagctttgctttgatcagatcc
ctattccactctggattagagaattacattttagtacttttcaaatatgtaatagataca
ctttttatctctatgtagattttaaactacataacaggactctttgtcatattgaatggt
ctgcagtattgctatctgaaattaccgataatattgtacattcagattcacttaagaggt
aaccttgcagagaat.....................................................

Exon 19
.....................................................cttaacaccatgccaggta
ccttagtaagtgttcgatgaatatttgctttttgtattagccataatcattctcaggctg
ctttgtcatttacttgttccacaaattcttagcttccaaaattttggtgatacctcattt
cctattctctctagttgccctttgtccatgtagatttttttgaggaagcttgggtaaataag
tgtattttaaactattatgtttaaatcgaagttccttttatctgttttctaatagaaaca
tttaaatagcattaagaacttgtagcagtataaacaatatgtttgagaagtactatattg
tgaaaatattttcacttttatacagtttttttacttatttactgtcttactaatcttccta
agacttttttaaagtgaatatttttaaggcagttctagaagaatgaaaactcttatgatat
ctgtaatagaattgaatacatatttaactactaaatcaatatatttattaatttgtccag

ATTTCTGCTAACAGTACTCGGCCTGCTCGCTGGTATACCAAACTTGGATTCTTTCCTGAC
I  S  A  N  S  T  R  P  A  R  W  Y  T  K  L  G  F  F  P  D   2797

CCTAGACCTTTTCCTCTGCCCTTATCATCGCTTTTCAGTGATGGAGGAAATGTTGGTTGT
P  R  P  F  P  L  P  L  S  S  L  F  S  D  G  G  N  V  G  C   2817

GTTGATGTAATTATTCAAAGAGCATACCCTATACAGgtatgatgtattcttgaaacttac
V  D  V  I  I  Q  R  A  Y  P  I  Q                           2829

catatatttctttcttttgatacaattaatttgtttgtttgtttgagatggagtttcggt
ctcttgcccaggctggagtgcaatggcgtgatcttggttcactgcagcctcccacctccc
gggttcaagtgattctcctgcctcagcctctcaagtagctgagccaccacacctggctaa
ttttgtatttttggtagagaaggggtttcatcatgttggtcaggctgatctcgaactcct
gacctcaggtgatccactaatctcagcctcccaaagttctgggattacagatgtgagcca
ctgtgcctggcctgatacaattaacttgaatgttatatatgtgacttttttggtgtgtgt
aacacattattacagtggatggagaagacatcatctggattatacatatttcgcaatgaa
agagaggaagaaaaggaagcagcaaaatatgtggaggcccaacaaaagagactaga....

Exon 20
.....................................ctgaccctagacctttttcct
ctgcccttatcatcgcttttcagtgatggaggaaatgttggttgtgttgatgtaattatt
```

Fig.7. (17/22)

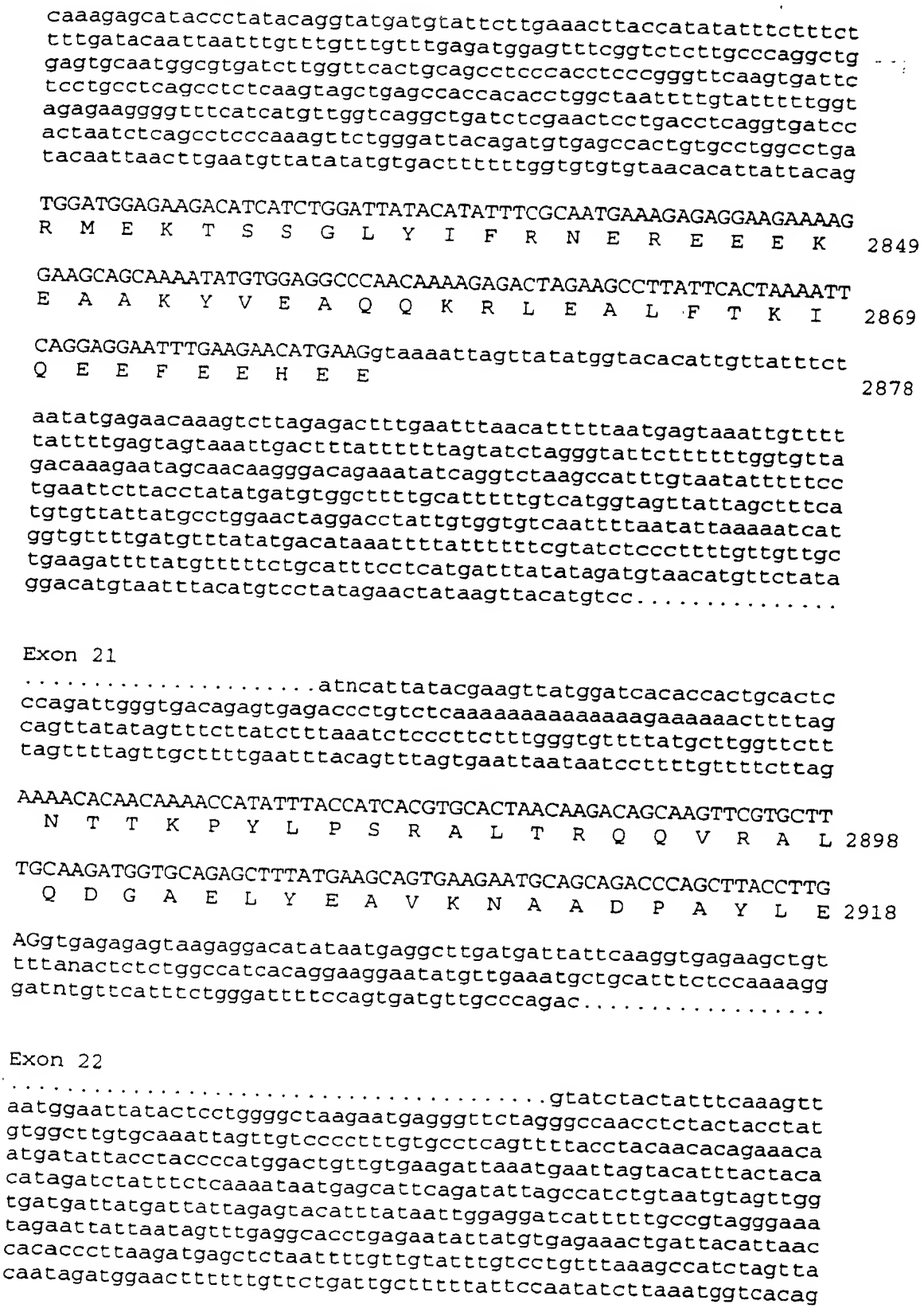

caaagagcataccctatacaggtatgatgtattcttgaaacttaccatatatttctttct
tttgatacaattaatttgtttgtttgtttgagatggagtttcggtctcttgcccaggctg
gagtgcaatggcgtgatcttggttcactgcagcctcccacctcccgggttcaagtgattc
tcctgcctcagcctctcaagtagctgagccaccacacctggctaattttgtatttttggt
agagaaggggtttcatcatgttggtcaggctgatctcgaactcctgacctcaggtgatcc
actaatctcagcctcccaaagttctgggattacagatgtgagccactgtgcctggcctga
tacaattaacttgaatgttatatatgtgacttttttggtgtgtgtaacacattattacag
TGGATGGAGAAGACATCATCTGGATTATACATATTTCGCAATGAAAGAGAGGAAGAAAAG
R M E K T S S G L Y I F R N E R E E E K 2849
GAAGCAGCAAAATATGTGGAGGCCCAACAAAAGAGACTAGAAGCCTTATTCACTAAAATT
E A A K Y V E A Q Q K R L E A L F T K I 2869
CAGGAGGAATTTGAAGAACATGAAGgtaaaattagttatatggtacacattgttatttct
Q E E F E E H E E 2878
aatatgagaacaaagtcttagagactttgaatttaacatttttaatgagtaaattgtttt
tattttgagtagtaaattgactttatttttttagtatctagggtattctttttttggtgtta
gacaaagaatagcaacaagggacagaaatatcaggtctaagccatttgtaatatttttcc
tgaattcttacctatatgatgtggcttttgcatttttgtcatggtagttattagctttca
tgtgttattatgcctggaactaggacctattgtggtgtcaattttaatattaaaaatcat
ggtgttttgatgtttatatgacataaattttatttttttcgtatctccctttttgttgttgc
tgaagattttatgtttttctgcatttcctcatgatttatatagatgtaacatgttctata
ggacatgtaatttacatgtcctatagaactataagttacatgtcc.....
Exon 21
.....atncattatacgaagttatggatcacaccactgcactc
ccagattgggtgacagagtgagaccctgtctcaaaaaaaaaaaaaagaaaaaacttttag
cagttatatagtttcttatctttaaatctcccttctttgggtgttttatgcttggttctt
tagttttagttgcttttgaatttacagtttagtgaattaataatccttttgttttcttag
AAAACACAACAAAACCATATTTACCATCACGTGCACTAACAAGACAGCAAGTTCGTGCTT
N T T K P Y L P S R A L T R Q Q V R A L 2898
TGCAAGATGGTGCAGAGCTTTATGAAGCAGTGAAGAATGCAGCAGACCCAGCTTACCTTG
Q D G A E L Y E A V K N A A D P A Y L E 2918
AGgtgagagagtaagaggacatataatgaggcttgatgattattcaaggtgagaagctgt
tttanactctctggccatcacaggaaggaatatgttgaaatgctgcatttctccaaaagg
gatntgttcatttctgggattttccagtgatgttgcccagac.....
Exon 22
.....gtatctactatttcaaagtt
aatggaattatactcctggggctaagaatgagggttctagggccaacctctactacctat
gtggcttgtgcaaattagttgtccccttttgtgcctcagttttacctacaacacagaaaca
atgatattacctaccccatggactgttgtgaagattaaatgaattagtacatttactaca
catagatctatttctcaaaataatgagcattcagatattagccatctgtaatgtagttgg
tgatgattatgattattagagtacatttataattggaggatcatttttgccgtagggaaa
tagaattattaatagtttgaggcacctgagaatattatgtgagaaactgattacattaac
cacacccttaagatgagctctaattttgttgtatttgtcctgtttaaagccatctagtta
caatagatggaacttttttgttctgattgctttttattccaatatcttaaatggtcacag

Fig.7. (18/22)

```
GGTTATTTCAGTGAAGAGCAGTTAAGAGCCTTGAATAATCACAGGCAAATGTTGAATGAT
G  Y  F  S  E  E  Q  L  R  A  L  N  N  H  R  Q  M  L  N  D   2938

AAGAAACAAGCTCAGATCCAGTTGGAAATTAGGAAGGCCATGGAATCTGCTGAACAAAAG
K  K  Q  A  Q  I  Q  L  E  I  R  K  A  M  E  S  A  E  Q  K   2958

GAACAAGGTTTATCAAGGGATGTCACAACCGTGTGGAAGTTGCGTATTGTAAGCTATTCA
E  Q  G  L  S  R  D  V  T  T  V  W  K  L  R  I  V  S  Y  S   2978

AAAAAAGAAAAAGATTCAGgtaagtatgtaaatgctttgtttttatcagttttattaact
K  K  E  K  D  S  V                                          2985

taaaaaatgaccttactaacaaaatgattataaatccagataaagtataaagttagttta
tatcagagaagcaaaatccactactaatgcccacaaagagataatataaaagaggatctg
tatttatttgaaacaaacatttaaatgataatcacttcttccattgcatctttctcatc
tttctccaaacagttatactgagtatttggcgtccatcatcagatttatattctctgtta
acagaaggaaagagatacagaatttatcatcttgcaacttcaaaatctaaaagtaaatct
gaaagagctaacatacagttagcagcgacaaaaaaaactcagtatcaacaactaccggta
caaacctttcattgtaatttttcagttttgataagtgcttgttagtttatggaatctcca
tatgttgaatttttgttttgttttctgtaggtttcagat.....................

Exon 23
.........................................ctagttacaatagatggaac
ttttttgttctgattgcttttattccaatatcttaaatggtcacagggttatttcagtg
aagagcagttaagagccttgaataatcacaggcaaatgttgaatgataagaaacaagctc
agatccagttggaaattaggaaggccatggaatctgctgaacaaaaggaacaaggtttat
caagggatgtcacaaccgtgtggaagttgcgtattgtaagctattcaaaaaaagaaaaag
attcaggtaagtatgtaaatgctttgtttttatcagttttattaacttaaaaaatgacct
tactaacaaaatgattataaatccagataaagtataaagttagtttatatcagagaagca
aaatccactactaatgcccacaaagagataatataaaagaggatctgtatttatttttgaa
acaaacatttaaatgataatcacttcttccattgcatctttctcatctttctccaaacag

TTATACTGAGTATTTGGCGTCCATCATCAGATTTATATTCTCTGTTAACAGAAGGAAAGA
  I  L  S  I  W  R  P  S  S  D  L  Y  S  L  L  T  E  G  K  R 3005

GATACAGAATTTATCATCTTGCAACTTCAAAATCTAAAAGTAAATCTGAAAGAGCTAACA
  Y  R  I  Y  H  L  A  T  S  K  S  K  S  K  S  E  R  A  N  I 3025

TACAGTTAGCAGCGACAAAAAAAACTCAGTATCAACAACTACCGgtacaaacctttcatt
  Q  L  A  A  T  K  K  T  Q  Y  Q  Q  L  P                   3039

gtaatttttcagttttgataagtgcttgttagtttatggaatctccatatgttgaatttt
tgttttgttttctgtaggtttcagatgaaatttttatttcagatttaccagccacgggagc
cccttcacttcagcaaatttttagatccagactttcagccatcttgttctgaggtggacc
taataggatttgtcgtttctgttgtgaaaaaaacaggtaatgcacaatatagttaatttt
ttttattgattctttttaaaaaacattgtcttttaaaatctcttatgattagttggagcta
ccagttggcaaatttgctagctaactagtgatctgaaagtaagcctctttgaacctctga
ttttttcatgaaaagcaattctctcaattctatattatttcaagggtaacaagttacatcc
tagtctgtgtacttaattttatagaaattgtccttaatttttatttttctgcaatttatgtt
ttct.....................................................
```

Fig.7. (19/22)

Exon 24

```
.....................................aagattcaggtaagtatgta
aatgctttgtttttatcagttttattaacttaaaaaatgaccttactaacaaaatgatta
taaatccagataaagtataaagttagtttatatcagagaagcaaaatccactactaatgc
ccacaaagagataatataaaagaggatctgtatttatttgaaacaaacatttaaatgat
aatcacttcttccattgcatctttctctcatctttctccaaacagttatactgagtatttgg
cgtccatcatcagatttatattctctgttaacagaaggaaagagatacagaatttatcat
cttgcaacttcaaaatctaaaagtaaatctgaaagagctaacatacagttagcagcgaca
aaaaaaactcagtatcaacaactaccggtacaaacctttcattgtaatttttcagttttg
ataagtgcttgttagtttatggaatctccatatgttgaatttttgttttgttttctgtag

GTTTCAGATGAAATTTTATTTCAGATTTACCAGCCACGGGAGCCCCTTCACTTCAGCAAA
V  S  D  E  I  L  F  Q  I  Y  Q  P  R  E  P  L  H  F  S  K   3059

TTTTTAGATCCAGACTTTCAGCCATCTTGTTCTGAGGTGGACCTAATAGGATTTGTCGTT
F  L  D  P  D  F  Q  P  S  C  S  E  V  D  L  I  G  F  V  V   3079

TCTGTTGTGAAAAAAACAGgtaatgcacaatatagttaatttttttttattgattctttta
S  V  V  K  K  T  G                                           3086

aaaaacattgtcttttaaaatctcttatgattagttggagctaccagttggcaaatttgc
tagctaactagtgatctgaaagtaagcctctttgaacctctgatttttcatgaaaagcaa
ttctctcaattctatattatttcaagggtaacaagttacatcctagtctgtgtacttaat
tttatagaaattgtccttaattttatttttctgcaatttatgttttcttactatttctggt
gtatgtgtttatcccattgtgatgttatattggtgtcctcaatttatttccttagccata
cactctacttttcattgtacagggctatttattatctcagagtcaagctttttttttttt
ttttttttccccgagatggagtctcactctgttgcccaggctggagtncagtggcgnaat
ntcagcccacngcaagttctgcctcccaggttnanacca.....................
```

Exon 25

```
.....................................tgtatatgagctaataaaag
ttgcttccaatattttataattataatgttttcagtgagtaaccttgttcataggtgttt
tcatgaactttatgttcatgtgtattttactattattagaggtctatcttcagagaggag
tacaagaaatgggattactgggtgcaaaggtaaatggatatgtgtctttgctaggtattg
ccaaatttatctccagaaatcttgcacaaatctgtactcctgttagcaatgtgtgcgtat
acctgcttaccacatgacctcagtaaaagaatgtgttgtcatattggtattgaaatttta
gcactgtaagcaacaggtcattttggaaaacctgagctttcgccaaattcagctattttg
atttgcttttattattagcatataccaaaataaataggcatattagagtttccttttcttg
catcttaaaattcatctaacacatctataataacattcttttctttttttttccattctag

GACTTGCCCCTTTCGTCTATTTGTCAGACGAATGTTACAATTTACTGGCAATAAAGTTTT
  L  A  P  F  V  Y  L  S  D  E  C  Y  N  L  L  A  I  K  F  W 3106

GGATAGACCTTAATGAGGACATTATTAAGCCTCATATGTTAATTGCTGCAAGCAACCTCC
  I  D  L  N  E  D  I  I  K  P  H  M  L  I  A  A  S  N  L  Q 3126

AGTGGCGACCAGAATCCAAATCAGGCCTTCTTACTTTATTTGCTGGAGATTTTTCTGTGT
  W  R  P  E  S  K  S  G  L  L  T  L  F  A  G  D  F  S  V  F 3146

TTTCTGCTAGTCCAAAAGAGGGCCACTTTCAAGAGACATTCAACAAAATGAAAAATACTG
  S  A  S  P  K  E  G  H  F  Q  E  T  F  N  K  M  K  N  T  V 3166

TTGAGgtaaggttacttttcagcatcaccacacattttggtatttttctattttgacagt
  E                                                           3167
```

Fig.7. (20/22)

```
ccagtatcaaggaaatagcttttatacaaattggatagttgaggtagtatgtgaggtaaa
gtttaatcatatattaattgcccatgaacctcaggagatgggggaatggggaaatgacag
cagctagaaagagaagaatgacttgaagggaaatgagttaggagaaattgtgagaaggat
gttcagaaatgcagactttgtaagcaaactggaaattggttacaagaataatatgagtta
tctgtggtttgcagcagtcagcagtgtgattagttaataatatagagactacaggtttac
atttaaactccatatctagtgttttatacagattatatttctttgacttgatttaatccc
agataagagacactgatattattttccctagatcatgt.....

Exon 26
.....tatttgaccagttgtgattt
ctcaagcaacagtttcatgtagtttaacctataaatcatttcaattaattcttggaacag
acgtgaggtaggtgaggcaattctttctttttctctaaccaaagaagtacctttatagatg
tgagatgattcccagctattaagtagtaaatagagctaggacttgagccccaatcttcca
gcttcaatccagatcatatgacagcttgctgattaaactagatgacagagaagatctctt
tccttcagatacacatacttttttctctgttcccctctccctatcagctagattcccctaa
atcactgatactggttttgtaattttgcatcggcatgtttgacaattggtatcacattta
gggtttttcattctttttttggtcccaaacttttcatttctgcttttaaaggaaatacttt
tggaaacataaatatgtgggtttgcaatttataaagcagcttttccacttatttttcttag

AATATTGACATACTTTGCAATGAAGCAGAAAACAAGCTTATGCATATACTGCATGCAAAT
N I D I L C N E A E N K L M H I L H A N  3187

GATCCCAAGTGGTCCACCCCAACTAAAGACTGTACTTCAGGGCCGTACACTGCTCAAATC
D P K W S T P T K D C T S G P Y T A Q I  3207

ATTCCTGGTACAGGAAACAAGCTTCTGgtaagttaatgtaaactcaaggaatattataag
I P G T G N K L L                      3216

aagtatatatggaggccatcgtatattctgttgtatacctagtaaacatggtaaaatgta
attaaacttaattagaaaatgtggttgttatgtggctcctgtaagtatagttatttagaa
attttatttattgaagcaagatatgaaactctgggtgcacactttccaaacaggtgcttt
catttacatgtgattgaaaagtgtttttttgccatttatttcactgttccatacaattagg
gttgtttctaagctgtttgtaagctgtttctaagctatttaagtggttaaatcacagtag
atgcaaatcaagctaaagtctttaacattggctaatggctgattcttaaatagctaatac
ttgctaagggtatctatattaactcatttaatcctcataacaaccctatgagataaaacc
taagtcctcacttaacattgtcaataggtttttggaaactgatttta.....

Exon 27
                                                    taatctgtt
aataataaaaaacaaaagattaaagcataagtgacgtcccctacctccttttttatcttt
tactgtgattattcttcatcttccttcctttcatgtcattttatatgttcttatgtaaa
attactttcatctagaataggaataatgtgaactgaaatcacctaacctattaggagtta
ggggagggagactgtgtgtaatatttgcgtgcttaaatattttcaatgaaaagttacttt
gatttagtttttttatgttactacataattatgataggctacgttttcatttttttatcag

ATGTCTTCTCCTAATTGTGAGATATATTATCAAAGTCCTTTATCACTTTGTATGGCCAAA
M S S P N C E I Y Y Q S P L S L C M A K  3236

AGGAAGTCTGTTTCCACACCTGTCTCAGCCCAGATGACTTCAAAGTCTTGTAAAGGGGAG
R K S V S T P V S A Q M T S K S C K G E  3256
```

Fig.7. (21/22)

```
AAAGAGATTGATGACCAAAAGAACTGCAAAAAGAGAAGAGCCTTGGATTTCTTGAGTAGA
K  E  I  D  D  Q  K  N  C  K  K  R  R  A  L  D  F  L  S  R   3276

CTGCCTTTACCTCCACCTGTTAGTCCCATTTGTACATTTGTTTCTCCGGCTGCACAGAAG
L  P  L  P  P  P  V  S  P  I  C  T  F  V  S  P  A  A  Q  K   3296

GCATTTCAGCCACCAAGGAGTTGTGGCACCAAATACGAAACACCCATAAAGAAAAAAGAA
A  F  Q  P  P  R  S  C  G  T  K  Y  E  T  P  I  K  K  K  E   3316

CTGAATTCTCCTCAGATGACTCCATTTAAAAAATTCAATGAAATTTCTCTTTTGGAAAGT
L  N  S  P  Q  M  T  P  F  K  K  F  N  E  I  S  L  L  E  S   3336

AATTCAATAGCTGACGAAGAACTTGCATTGATAAATACCCAAGCTCTTTTGTCTGGTTCA
N  S  I  A  D  E  E  L  A  L  I  N  T  Q  A  L  L  S  G  S   3356

ACAGGAGAAAAACAATTTATATCTGTCAGTGAATCCACTAGGACTGCTCCCACCAGTTCA
T  G  E  K  Q  F  I  S  V  S  E  S  T  R  T  A  P  T  S  S   3376

GAAGATTATCTCAGACTGAAACGACGTTGTACTACATCTCTGATCAAAGAACAGGAGAGT
E  D  Y  L  R  L  K  R  R  C  T  T  S  L  I  K  E  Q  E  S   3396

TCCCAGGCCAGTACGGAAGAATGTGAGAAAAATAAGCAGGACACAATTACAACTAAAAAA
S  Q  A  S  T  E  E  C  E  K  N  K  Q  D  T  I  T  T  K  K   3416

TATATCTAAGCATTTGCAAAGGCGACAATAAATTATTGACGCTTAACCTTTCCAGTTTAT
Y  I  STOP                                                    3418

AAGACTGGAATATAATTTCAAACCACACATTAGTACTTATGTTGCACAATGAGCAAAGAA

ATTAGTTTCAAATTTACCTCAGCGTTTGTGTATCGGGCAAAAATCGTTTTGCCCGATTCC

GTATTGGTATACTTTTGCTTCAGTTGCATATCTTAAAACTAAATGTAATTTATTAACTAA

TCAAGAAAAACATCTTTGGCTGAGCTCGGTGGCTCATGCCTGTAATCCCAACACTTTGAG

AAGCTGAGGTGGGAGGAGTGCTTGAGGCCAGGAGTTCAAGACCAGCCTGGGCAACATAGG

GAGACCCCCATCTTTACAAAGAAAAAAAAAAGGGGAAAAGAAAATCTTTTAAATCTTTGG

ATTTGATCACTACAAGTATTATTTTACAAGTGAAATAAACATACCATTTTCTTTTAGATT

GTGTCATTAAATGGAATGAGGTCTCTTAGTACAGTTATTTTGATGCAGATAATTCCTTTT

AGTTTAGCTACTATTTTAGGGGATTTTTTTTAGAGGTAACTCACTATGAAATAGTTCTCC

TTAATGCAAATATGTTGGTTCTGCTATAGTTCCATCCTGTTCAAAAGTCAGGATGAATAT

GAAGAGTGGTGTTTCCTTTTGAGCAATTCTTCATCCTTAAGTCAGCATGATTATAAGAAA

AATAGAACCCTCAGTGTAACTCTAATTCCTTTTTACTATTCCAGTGTGATCTCTGAAATT

AAATTACTTCAACTAAAAATTCAAATACTTTAAATCAGAAGATTTCATAGTTAATTTATT
```

Fig.7. (22/22)

TTTTTTTTcaacaaaatggtcatccaaactcaaacttgagaaaatatcttgctttcaaat

tggcactgattctgcctgctttatttttagcgctatcacaggacccagagcctatgccct
tttaaacttaccacaaaagcagaagattaattcaatttaagatgatactctcatttgtta
cgtccttttttttttttttggagatggagtcttgctttgtcgcccatgctggagtgcag
tggcatgatcctggctcactgcagcctccacttcccgggttcaagtaattctcccacctc
aagcctccctagtagctgggattacagggacgcaccaccatgcccagctaatttttgcat
ttttagtagagactgggttttaccatgttggccaagctggtctcaaactcctgatgtcag
gtgatccatctgcctcagcctcccaaagtgctgggattataggcgtgagccactgtgccc
ggccaatatttgttactttcttaggtttaatagagaaaagggataaaacatttctaactg
ggagttaatagcatggagaaggtcttaaatcagatgttttaatgccttaaatgtctgtat
aatatcatgttttcaaatctaattataaatacgtttaaagccaagaataaatcttttaaa
aaattgacttgtttccttccataactctgagccatgatttttctgttctgtaaaaagcat
taacaaaattgtctattttgctactccctgtaacttaagtattctgcaagtcttattaat
gagacttgttttgtttctaaaacagtttggttttcacatcctaattttgcagtgatccac
tctagaacaaggaaataaaacttgggtttcaaacaggagaacagaaaaaattacaagaat
ttaacctttttctttttgaatccttggtacaactgctattgtctgttctcatgtagaaca
cccattatgttgatagatatgtaatgcgcacacttttttataaattataaataacatcca
agctacatgaaacaaaatatgaagcttgagtatatgtgcatattatcccctcaaaagtga
caatttaatgactacaaagtcaacatactgcatcataagggataatggtaaaaatttgtg
ttatttattcagcaatcatttaatgagaccctattgcccataaagagcatttgccagttc
tatgaatgatgcaaacagcgaacacaatacaagtcaatattgggtgttcaaagagttaca
caaatcagaagccatgggagttgaaaaacagatcacttctacaggggataaagttgatgg
catttaatcaatttttatggtttacaaaccaaatatttcaggggtcagctagggaaagaa
gagacattaaataggccaggaaactattctttagagctctagaaaagttatgtaacccag
agccacttctccatagtggtcactcacttaacccaagtgagcca.....

Fig.8. (1/4)

BRCA2 SSCP PRIMERS

EXON	PRIMER F	PRIMER R	SIZE(bp)	CON
1				
2	CTCAGTCACATAATAAGGAAT	ACACTGTGACGTACTGGGTTTT	256	TD60-50
3	TCTGGGTCACAAATTTGTCTGTCA	TTCCTAGTTTGTAGTTCTCCCCAGTC	356	TD60-50
4	AGAATGCAAATTTATAATCCAGAGTA	AATCAGATTCATCTTTATAGAACAAA	249	TD60-50
5	AACAATTTATATGAATGAGAATC	AATTGTTAAGTTTTATTTTTATTA	220	TD55-45
6	CCACAAAGAGATAAGTCAGGTA	TGTAAATCTCAGGGCAAAGGTA	234	TD55-45
7	TAAGTGAAATAAAGAGTGAA	AACAGAAGTATTAGAGATGAC	275	TD55-45
8	AATAGTAGATGTGCTTTTTGA	ACATATAGGACCAGGTTTAGAGAC	285	TD60-50
9	GTTCAACTAAACAGAGGACT	CTAGTGATTTTAAACTATAATTTTTG	164	TD68-60
10A	TATAAAATATTAATGTGCTTCTGTT	AAAGGGCTTCTGATTTGCTAC	374	TD55-45

Fig.8. (2/4)

10B	ATCTGAAGTGGAACCAAATGATAC	ACGTGGCAAAGAATTCTCTGAAGTAA	270	TD60-50
10C	TTTCAGAAAAAGACCTATTAGACA	CTTTTTGATACCCTGAAATGAAGAAG	260	TD60-50
10D	TAAAGCAGGCAATATCTGGAACTTCT	GTGGATATTAAACCTGCATTCTTCAA	280	TD60-50
10E	TATGTCCAAATTTAATTGATAAT	AAACACAGAAGGAATCGTCATC	320	TD55-45
11A	GATGGTACTTTAATTTTGTCACTTTG	TTCCTTATTACATTTTGCTTCTTTAT	220	TD55-45
11B	CTAGCTCTTTTGGGACAATTCTGAGG	TTTCATGATCATATAAAAGANT	300	TD55-45
11C	CAAAAGTGGAATACAGTGATAC	ATAATTTTCATTTAAAGCACATACAT	250	TD60-50
11D	TCTAGAGGCAAAGAATCATA	TCATTGTCTGAGAAAAGTTC	305	TD55-45
11E	TTCAAAAATAACTGTCAATCC	GTTGCTTGTTTATCACCTGT	230	TD60-50
11F	AACCCATTTTCAAGAACTCTACCA	CTGAAGCTACCTCCAAAACTGTG	274	TD60-50
11G	ACAAATGGGCAGGACTCTTAGG	CCTGCTTGGAAAATAACATCTG	300	TD60-50
11H	AGTTGTTTCTGATTGTAAAAATAGTC	GCTGCTGTCTACCTGACCAA	310	TD60-50
11I	GATGCTGATCTTCATGTCATAA	TACCTCTGCAGAAGTTTCCTCACTA	300	TD60-50
11J	GGGTTTAGGGGCTTTTATTC	TATCAGTTGGCATTTATTATTTTT	237	TD60-50
11K	CTTCAAGTAAATGTCATGATTCTGTC	CTGGCAGCAGTATATTTGTTATCT	202	TD60-50
11L	ATTACTGAAAATTACAAGAGAAATA	AAAAAGTTAAATCTGACTCAAATCT	244	TD60-50
11M	TTTATGAAGGAGGGAAACACTCA	TTCAGAATTTAAGGAAAAGTTATGC	282	TD60-50
11N	AAAAATATTAGTGTCGCCAAAGAG	TGTATGAAAACCCAACAGAGTAGG	278	TD60-50
11O	GTTAAACACAAAATACTGAAAG	CATTGATGGCTAAAACTGGTG	233	TD60-50
11P	TCATACAGCTAGCGGGAAAAA	TGGCACCACAGTCTCAATAG	265	TD60-50
11Q	CTGCCCCAAAGTGTAAAGAAAT	AATGACTGAATAAGGGGACTGAT	230	TD60-50
11R	GTCCTGCAACTTGTTACAC	GATTTTTGTCATTTTCAGC	230	TD60-50
11S	AACCAGAAAGAATAAATACT	TCCTCAACGCAAATATCTTCAT	334	TD55-45
11T	TTTCCAAAGTAATATCCAATGTA	ATTTTTGATTTATTCTCGTTGTT	291	TD55-45
11U	AAGACATATTTACAGACAGT	TGAAGCTTCCCTATACTACAT	300	TD60-50
11V	CACCTTGTGATGTTAGTTTG	TTGGGATATTAAATGTTCTGGAGTA	350	TD60-50
11W	AAAGTAACGAACATTCAGACCA	CTGGGTTTCTCTTATCAACNCGA	310	TD60-50
11X	AGTCTTCACTATTCACCTACG	GTGAGACTTTGGTTCCTAAT	280	TD60-50
11Y	TTCAACAAGACAAACAACAGT	TGTCAGTTCATCATCTTCCATAAA	270	TD60-50
11Z	TTACTCCAAAGATTCAGAAAACTAC	AGCATACCAAGTCTACTGAATAAAC	290	TD60-50

Fig.8. (3/4)

12	AGGTCACTATTTGTTGTAAG	AGTGGCTCATGTCTGTAAT	358	TD60-50
13	TAAAGCCTATAATTGTCTCA	CTTCTTAACGTTAGTGTCATT	271	TD55-45
14A	ATGTAGCAAATGAGGGTCTG	GTCTGCCTGTAGTAATCAAGTGT	289	TD60-50
14B	CTTCAAGCAATTTAGCAGTTTCAG	TGCTGCTTGATTGGAGTTGTT	292	TD55-45
14C	TATTAACTTGGAGGAAAACAGACA	CAAAGGGGGAAAACCATCAG	250	TD60-50
15	GGCCAGGGGTTGTGCTTTTT	AGGATACTAGTTAATGAAATA	314	TD60-50
16	TTTGGTAAATTCAGTTTTGGTTT	AACACACAATCTTTTTGCATAGA	330	TD60-50
17	CAGAGAATAGTTGTAGTTGTTGAA	AGAAACCTTAACCCATACTGC	306	TD60-50
18A	ATTCAGTTTTTATTCTCAGTTATTC	ACCACCCATCTGTAAGTTCA	296	TD60-50
18B	ATCTGAAACTTCTAGCAATAA	TTTAACTGAATCAATGACTG	281	TD60-50
19	AAGTGAATATTTTTAAGGCAGTT	TATATGGTAAGTTTCAAGAAT	296	TD60-50
20	CACTGTGCCTGGCCTGATAC	ATGTTAAATTCAAAGTCTCTA	296	TD60-50
21	GGGTGTTTTATGCTTGGTTCT	CATTTCAACATATTCCTTCCTG	304	TD60-50
22	TTTTGTTCTGATTGCTTTTTATTC	AATCATTTTGTTAGTAAGGTCAT	314	TD60-50
23	ACTTCTTCCATTGCATCTTTCTCA	AAAACAAAACAAAAATTCAACATA	290	TD60-50
24	CAGTTTTGATAAGTGCTTGTT	AGCTCCAACTAATCATAAGA	290	TD60-50
25A	TAAAATTCATCTAACACATCTAT	ACCTCAACAGTATTTTTCATT	302	TD55-45
25B	AATCCCAAATCAGGCCTTCTTACTT	ATTTCCCCATTCCCCCATCT	283	TD60-50

Fig.8. (4/4)

26	AAGGAAATACTTTTGGAAACATAA	TTTACTAGGTATACAACAGAA	299	TD55-45
27A	TAGGAGTTAGGGGAGGGAGACTGTGT	TTTCTCCCCTTTACAAGACTTTGAAG	252	TD60-50
27B	TTGTATGGCCAAAAGGAAGTCTG	TTTTTAAATGGAGTCATCTGAGGAGA	285	TD60-50
27C	AAGGCATTTCAGCCACCAAGGAGT	CAACGTCGTTTCAGTCTGAGATAATC	272	TD60-50
27D	AGGAGAAAAACAATTTATATCTGTCA	GTGGTTTGAAATTATATTCCAGTCTT	264	TD60-50

Fig.9.

BRCA2 PTT primers

Exon	Primers	PTT primer sequence	Size
10	JAD20F CAAAGACCACATTGGAAAGTC x JAD20R GCAAATGTAAGTGGTGCTTC	PTTH2 GAAACAGTTGTAGATACCTCTGAAGA	1kb approx. 37kd
11	JAD53F GATGGTACTTTAATTTTGTCAC x JAD13R AAATGACTCTTTGGCGACAC	PTTD2 GATTCTGAAGAACCAACTTTGTCC	2.4kb approx. 89kd
11	JAD12F AGATTTTGAGACTTCTGATAC x JAD1R AGCATACCAAGTCTACTGAATAAAC	PTTA GACATTCTAAGTTATGAGGA	2.4kb approx. 89kd
27	JAD43F ACATAATTATGATAGGCTACG x JAD45R AAGCGTCAATAATTTATTGTC	PTTJ TCTTCTCCTAATTGTGAGATA	0.6kb approx. 22kd

Fig.10. (1/4)

4952

Fig.10. (2/4)

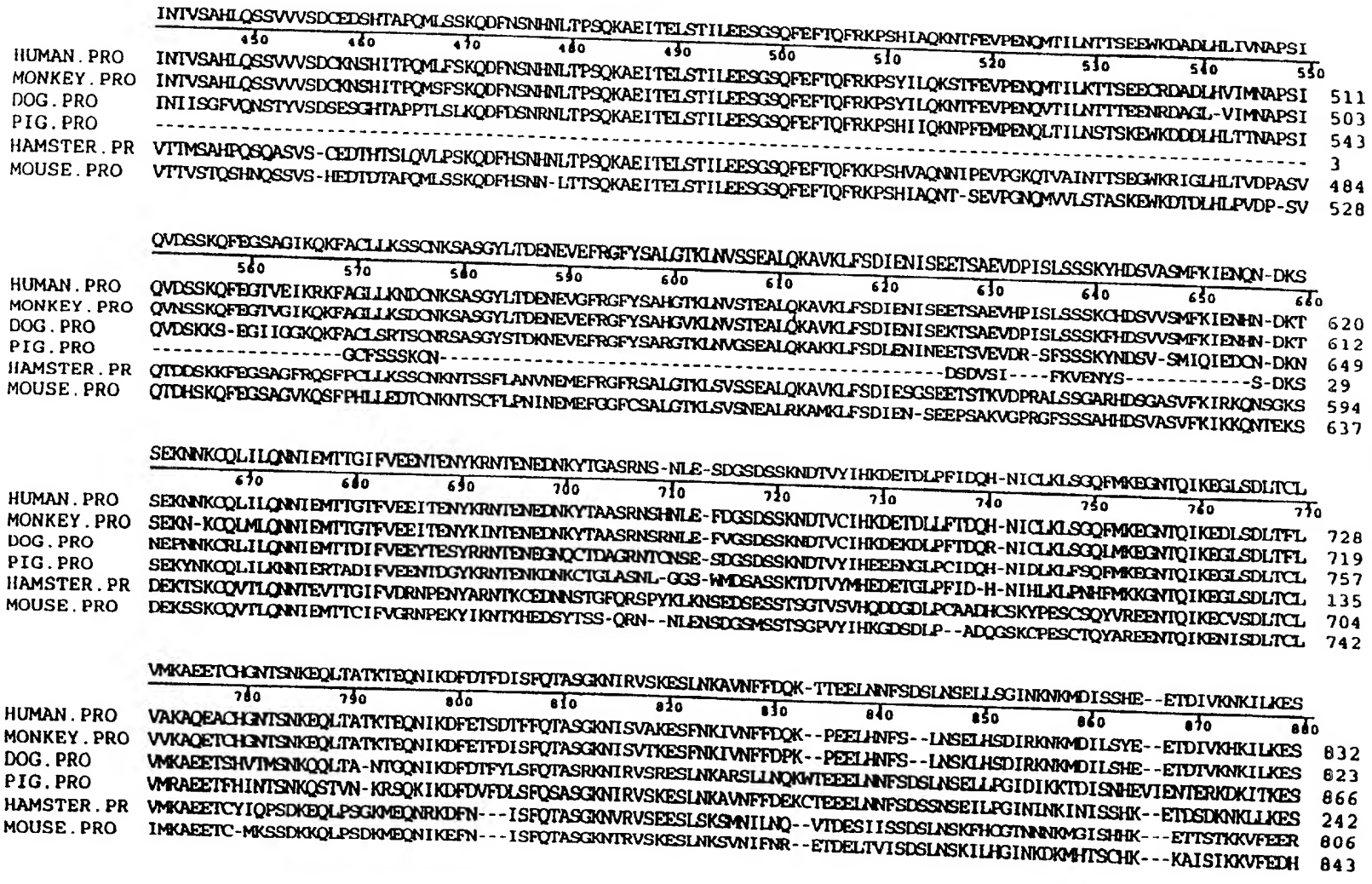


Fig.10. (3/4)

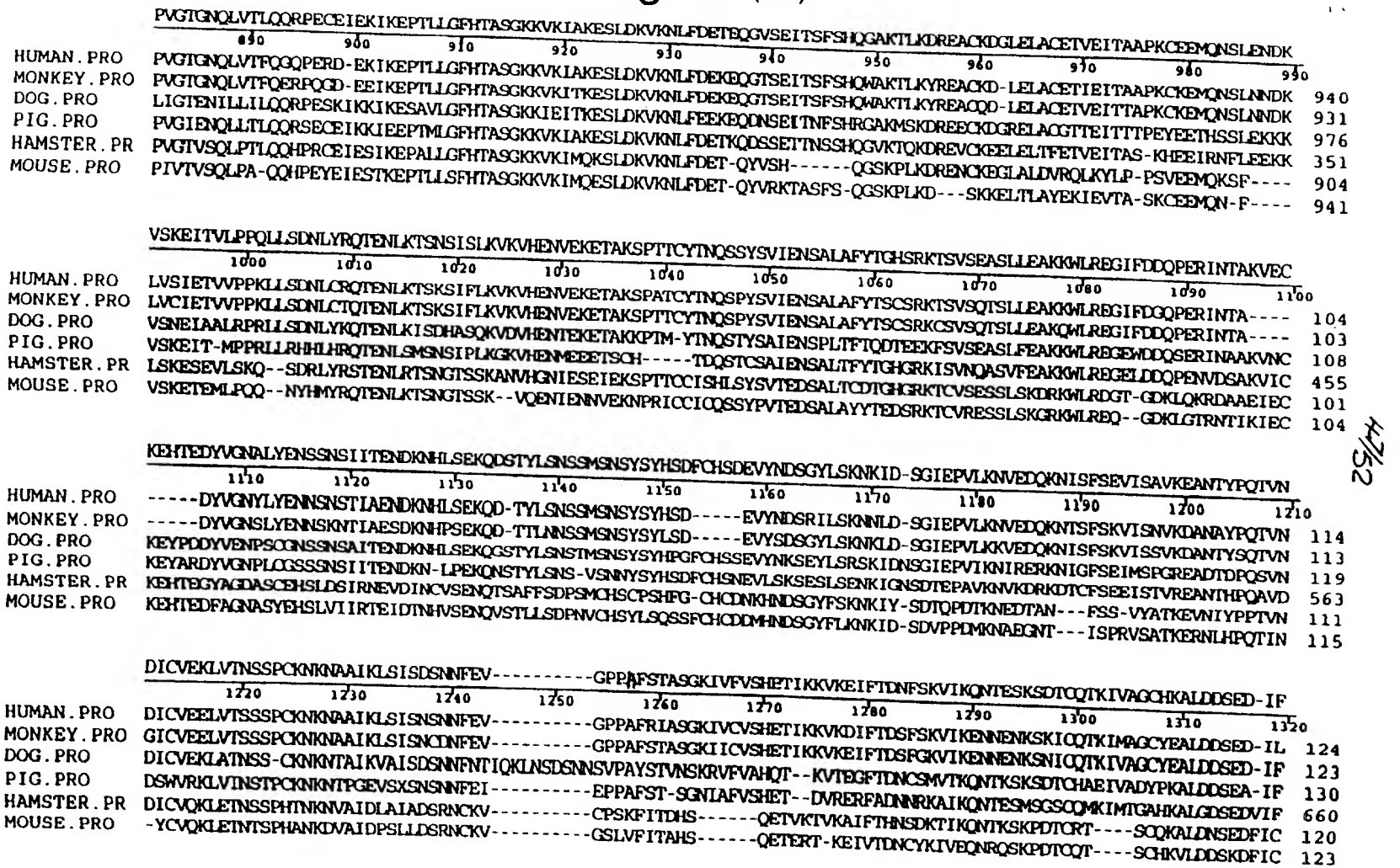


Fig.10. (4/4)

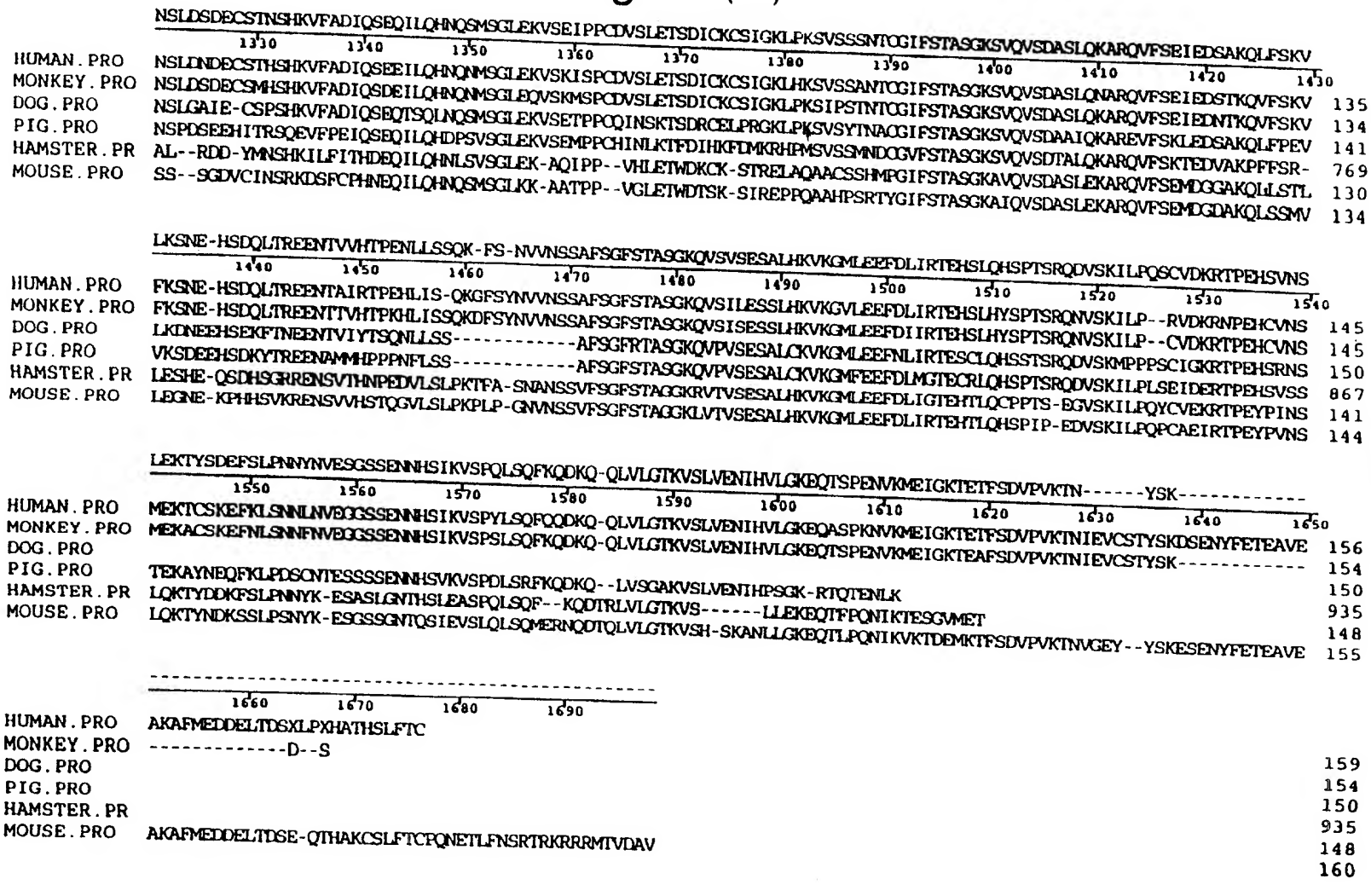


Fig.11.

Fig.12.

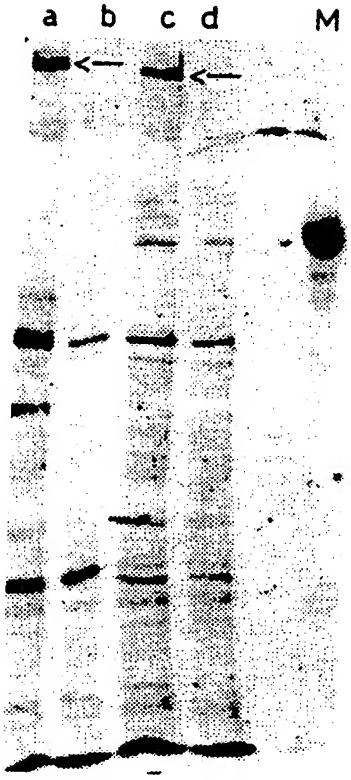


Fig.13A.

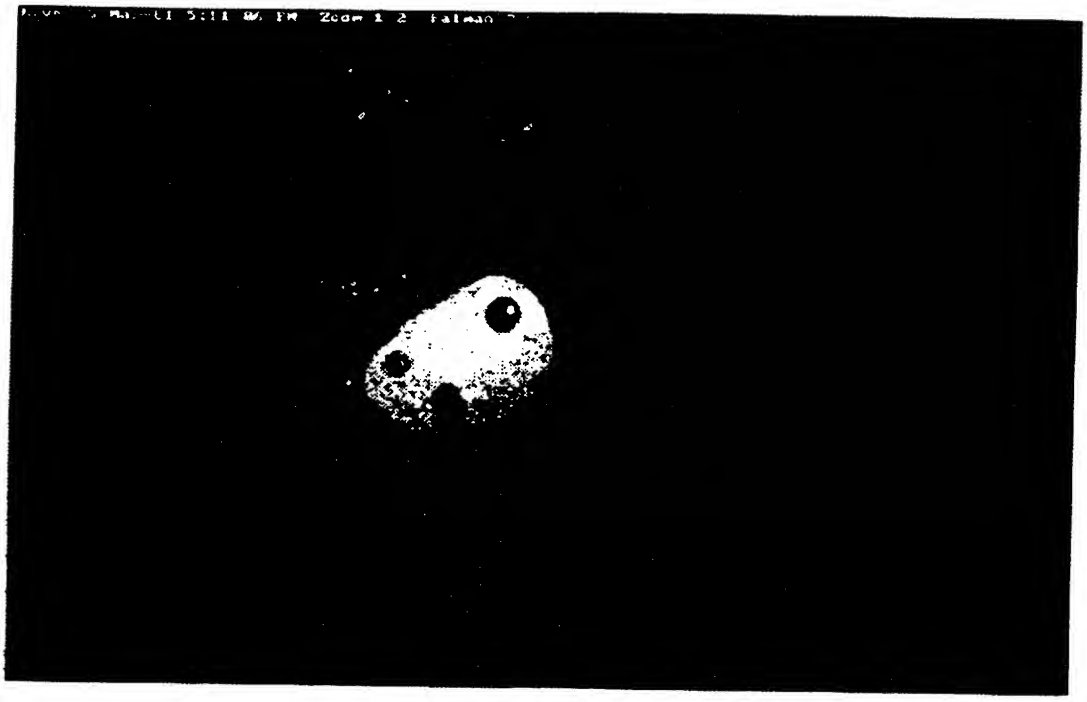
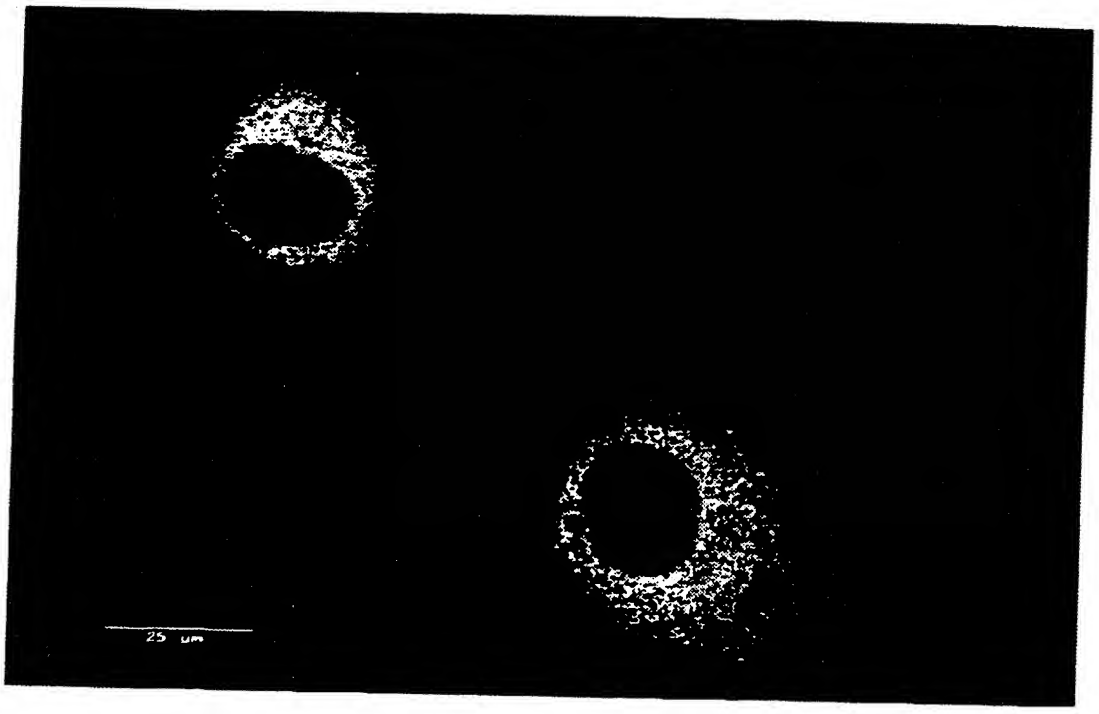


Fig.13B.



Materials and Methods Relating to the Identification and Sequencing of the BRCA2 Cancer Susceptibility Gene and Uses Thereof

Field of the Invention

The present invention relates to the identification and sequencing of the BRCA2 cancer susceptibility gene, and to materials and methods deriving from these findings. In particular, the present invention relates to nucleic acid molecules encoding BRCA2 polypeptides, and alleles of the BRCA2 gene, including those with mutations which are associated with a predisposition to develop cancer, especially breast and ovarian cancer, and to polypeptides encoded by this nucleic acid. The present invention further relates to uses of such BRCA2 nucleic acid and BRCA2 polypeptides, in particular in the diagnostic, prognostic or therapeutic treatment of cancer.

Background of the Invention

Over a lifetime approximately 1 in 12 women develops cancer of the breast. While a large majority of these cancers are thought to be sporadic, a proportion of breast cancer cases, often quoted at approximately 5%, is attributable to a predisposition to the disease which is transmitted as a highly penetrant autosomal dominant trait. This usually manifests as familial clustering of early onset breast cancer cases which is often associated with cancers of other organs, notably the ovary.

Abnormalities of several genes are known to confer susceptibility to breast cancer. The BRCA1 gene is located on chromosome 17q21 (1). BRCA1 encodes a 1863 aa protein which contains a RING finger domain and has little other homology to previously characterised proteins (3). Germline mutations of BRCA1 usually result in truncation or absence of the protein and hence presumed inactivation of one or more of its critical functions. BRCA1 accounts for approximately a third of families with site specific breast cancer. A small proportion of familial breast cancers are attributable to germline mutations in the p53 gene and rare clusters of male breast cancers (4) are due to mutations in the androgen receptor. Work relating to the BRCA1 gene, methods used to isolate it and applications of the BRCA1 nucleic acid and polypeptides are disclosed in EP-A-0705902.

Using families with multiple cases of early onset breast cancer showing evidence against linkage to BRCA1, the present inventors recently demonstrated the existence of a second major breast cancer susceptibility locus, BRCA2, on chromosome 13q12-q13 (5). Preliminary studies indicate that mutations in BRCA2 confer a similar risk of breast cancer to BRCA1. However, the risk of ovarian cancer appears to

be lower and the risk of male breast cancer substantially higher. Together BRCA1 and BRCA2 account for three quarters of families with multiple early onset breast cancer cases, and almost all families with both breast and ovarian cancer.

Summary of the Invention

Broadly, the work described in this application is based on the identification of the BRCA2 gene, and the disclosure of its sequence and the amino acid sequence of the corresponding BRCA2 polypeptides. BRCA2 alleles, including those with mutations in the normal sequence, are also disclosed.

The inventors initially found a portion of exon 16 of the BRCA2 gene and used this to sequence approximately 10% of the coding sequence. The BRCA2 cDNA sequence corresponding to this portion of the BRCA2 gene is shown in figure 1, the genomic sequence of the introns and exons initially sequenced by the inventors is shown in figure 2, with the translated protein sequence being shown in figure 3.

Just after the identification of 10% of the coding sequence of the BRCA2 gene by the present inventors, a 900,000bp sequence was released on the Internet on 23/11/95 and can be accessed at ftp://ftp.sanger.ac.uk/pub/human/sequences/13q and on ftp://genome/wustl.edu/pub/gscl/brca2.

Further work by the present inventors using this sequence and other known methods isolated around 75% of the BRCA2 coding sequence. This is shown in the cDNA sequence of figure 4, with the corresponding translated amino acid sequence shown in figure 5. The inventors also at this time identified 6 mutations in the BRCA2 gene, most notably a mutation at 6174delT from analysis of pedigrees from Ashkenazi Jewish families from Montreal.

The full BRCA2 sequence, showing the intron and exon structure is set out in figure 7, with the promoter region of the BRCA2 locus shown in figure 7. Primers suitable for amplifying the BRCA2 sequence are shown in figure 8 and are also published on 12/3/1996 in Nature Genetics, 12:333-337, 1996.

Following the initial sequencing of the BRCA2 gene described above, and using the information contained in figures 1 to 3, the skilled person could readily assemble the full length sequence of the BRCA2 gene included in the Internet sequence using the techniques described in detail below.

In a first aspect, the present invention provides a nucleic acid molecule comprising a part of the BRCA2 gene as set out in figures 1 or 2, or alleles thereof.

This nucleic acid can be used to obtain the full length sequence of the BRCA2 gene. Accordingly, in a further aspect, the present invention provides a nucleic acid molecule comprising the full length coding sequence or complete BRCA2 gene as obtainable by:

(a) using the nucleic acid sequences shown in figures 1 or 2 to construct probes for screening cDNA or genomic libraries, sequencing the positive clones obtained and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained;

(b) using the sequences shown in figures 1 or 2 to obtain oligonucleotides for priming BRCA2 nucleic acid fragments, these oligonucleotides being used in conjunction with oligonucleotides designed to prime from a cloning vector, to amplify by PCR nucleic acid fragments in a library that contains fragments of the BRCA2 sequence, sequencing the amplified fragments to obtain the BRCA2 sequence between known parts of the sequence and the cloning vector, and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained; and/or,

(c) using rapid amplification of cDNA ends (RACE), by synthesizing cDNAs from a number of different tissue RNAs, the cDNAs being ligated to an oligonucleotide linker, and amplifying by PCR the BRCA2 cDNAs using one primer that primes from the BRCA2 cDNA sequence of figure 1 and a second primer that primes from the oligonucleotide linker, sequencing the amplified nucleic acid and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained.

In a further aspect, the present invention provides a nucleic acid molecule comprising a part of the BRCA2 gene as set out in figure 4, or alleles thereof. The nucleic acid of figure 4 can be used in the same way as the nucleic acid set out in figures 1 and 2 to obtain the coding sequence of the full length BRCA2 gene.

The sequences set out in figures 1, 2 and 4 are believed to be a rare alternative splice of BRCA2 including nucleic acid at the 3' end of exon 16 coding for an additional 8 amino acids (ALCDVKAT). The sequence in figure 7 shows what is thought to be the normal amino acid sequence of the BRCA2 polypeptide at this position. However, the presence of the alternative splice has no effect on the methodology outlined above for isolating the full length BRCA2 gene using the 10% and 75% sequences initially isolated by the present inventors.

In a further aspect, the present invention provides a nucleic acid molecule which has a nucleotide sequence encoding a BRCA2 polypeptide including the amino acid sequence set out in any one of figures 3, 5 or 7.

In a further aspect, the present invention provides a nucleic acid molecule which has a nucleotide sequence encoding a polypeptide which is an allele (including mutant alleles) or variant of a BRCA2 polypeptide including the amino acid sequence set out in any one of figure 3, 5 or 7. Where the nucleic acid sequence is a mutant allele sequence, preferably it includes one or more of the exemplary mutations set out in table 1. Preferred mutations from table 1 include the 6174delT and 6503delTT mutations.

In a further aspect, the present invention provides a nucleic acid molecule which has a nucleotide sequence encoding a fragment or active portion of a BRCA2 polypeptide including the amino acid sequence set out in any one of figure 3, 5, or 7.

In a further aspect, the present invention provides nucleic acid encoding all or a part the BRCA2 promoter region, the nucleic acid sequence of which is set out in figure 6.

In further aspects, the present invention includes replicable vectors comprising the above nucleic acid operably linked to control sequences to direct its expression, host cells transformed with these vectors, and methods of producing BRCA2 polypeptide comprising culturing the host cells and recovering the polypeptide produced.

In a further aspect, the present invention provides the above nucleic acid molecules for use in methods of medical treatment, especially in the diagnosis and therapy of cancer. Also included herein is the use of the above nucleic acid molecules in the preparation of a medicament for treating cancer. This is discussed further below.

In a further aspect, the present invention provides the use of one of the above nucleic acid sequences in the design of primers for use in the polymerase chain reaction.

In a further aspect, the present invention provides substances comprising polypeptides encoded by the above nucleic acid, or an active portions, derivatives or functional mimetics thereof.

In a further aspect, the present invention provides a method of diagnosing a susceptibility or predisposition to cancer in a patient by

analysing a sample from the patient for the BRCA2 gene or the polypeptide encoded by it. By way of example, this could be carried out by:

(a) comparing the sequence of nucleic acid in the sample with the BRCA2 nucleic acid sequence to determine whether the sample from the patient contains mutations; or,

(b) determining the presence in a sample from a patient of the polypeptide encoded by the BRCA2 gene as set out in the partial sequences of figures 3 and 5 or the full length sequence set out in figure 7 and, if present, determining whether the polypeptide is full length, and/or is mutated, and/or is expressed at the normal level; or,

(c) using DNA fingerprinting to compare the restriction pattern produced when a restriction enzyme cuts a sample of nucleic acid from the patient with the restriction pattern obtained from normal BRCA2 gene comprising the sequence set out in figures 1, 2, 4 or 7 or from known mutations thereof; or,

(d) using a specific binding member capable of binding to a BRCA2 nucleic acid sequence (either a normal sequence or a known mutated sequence), the specific binding member comprising nucleic acid hybridisable with the BRCA2 sequence, or substances comprising an antibody domain with specificity for a native or mutated BRCA2 nucleic acid sequence or the polypeptide encoded by it, and detecting the binding of the specific binding member to its binding partner by means of a label; or,

(e) using PCR involving one or more primers based on normal or mutated BRCA2 gene sequence to screen for normal or mutant BRCA2 gene in a sample from a patient.

While diagnostic methods (a)-(e) are provided as examples, other assay formats are well known in the art and will be apparent to the skilled person.

The detection of mutations in the BRCA2 gene indicates a susceptibility to cancer, especially female breast cancer, male breast cancer and ovarian cancer. Risks of other cancers including prostate cancer, pancreatic cancer, ocular melanoma, colorectal cancer and leukaemia are also likely to be elevated in carriers of BRCA2 mutations.

Brief Description of the Drawings

The above and further aspects of the present invention will now be further described by way of example with reference to the accompanying drawings, by way of example and not limitation. Still further aspects of the invention will be apparent to those or ordinary skill in the art.

Figure 1 shows the partial sequence of the BRCA2 gene obtained from a cDNA clone 14.

Figure 2(a)-(e) shows the sequences of the exons and introns of clone 14 where they are known.

Figure 3 shows the translated amino acid sequence of nucleic acid sequence shown in figure 1. Of the 3 possible reading frames, the one used to translate the amino acid sequence shown in this figure is the most likely candidate as the nucleic acid obtained when the reading frame is in the other positions contains stop codons in the sequence.

Figure 4 shows the second part of the cDNA sequence of the BRCA2 gene isolated by the inventors, including the cDNA sequence set out in figure 1.

Figure 5(a)-(d) shows the translated amino acid sequence of nucleic acid sequence shown in figure 4, with the bold arrows over the sequence indicating the boundaries between the parts of the protein encoded by different exons in the gene.

Figure 6 shows the sequences of the BRCA2 promoter region with the binding sites for potential transcription factors underlined and CpG shown in bold. The promoter sequence may be of use in screening for substances which modulate the expression of the BRCA2 gene for use a therapeutics.

Figure 7 shows the sequence of the BRCA2 gene, including the exon/intron structure defined by comparing the BRCA2 cDNA sequence with the genomic sequence of chromosome 13q between D13S260 and D13S171. Exons are shown in upper case with the flanking intron sequence shown in lower case. The amino acid sequence is shown below the open reading frame from exon 2 to 26.

Figure 8 shows primers for single stranded conformation polymorphism (SSCP) testing which have been designed to amplify the BRCA2 gene by the PCR for the identification of sequence changes in genomic DNA that may predispose to the development of breast cancer. The primers are located in the intron sequences flanking each of the 26 exons that contain the open reading frame of the gene. The amplified products include the splice site consensus sequences. Two or more sets of primers have been designed for the larger exons of the gene. The primers are labelled by: their exon number, the subsection of the exon and as forward or reverse. The PCR products range from 160 to 360bp. The column labelled CON indicates the conditions that we have used.

The primers have been tested using a limited set of conditions which were all based on touchdown (TD) PCR where the figures indicate the first and final annealing temperature respectively. When mutations have been identified, the PCR products have been sequenced using the same primers and fluorescent cycle sequencing.

Figure 9 shows the alignment of amino acid motifs in exon 11 of BRCA2 from different species.

Figure 10 shows exemplary primers for use in PTT assays.

Figure 11 shows immunoprecipitation of FLAG-tagged BRCA2 protein from COS cells transfected with expression plasmid p13120. Total cell lysates from transfected, (lanes a and b) and untransfected (lanes c and d) COS cells subjected to immunoprecipitation with anti-FLAG antibody (lanes b and d) or with a control antibody against a Golgi protein (lanes a and c). Anti-FLAG antibody immunoprecipitates a protein of approximately 400kDa (arrow) from transfected, but not from untransfected COS cell lysate (compare lanes d and b). The control antibody does not immunoprecipitate this protein (see lane c). The strong, high molecular weight protein in the marker lanes is 220kDa.

Figure 12 shows a Western blot of FLAG-tagged BRCA2 protein. Total cell lysates of COS cells (lanes a and b) and 293T cells (lanes c and d) transfected with expression plasmid p13120 (lanes a and c) or no DNA (lanes b and d). Anti-FLAG antibody detects a protein of approximately 400kDa (arrow) in the transfected, but not in the untransfected cell lysates. The protein in the marker lane has a molecular weight of 220kDa.

Figure 13 shows the detection of FLAG-tagged BRCA2 protein by immunofluorescence. COS cells transfected with expression plasmid, p13120 and fixed, 48 hours later, with 4% formaldehyde, then permeablised with 0.4% Triton-X100. Anti-FLAG antibody reveals variation in the cellular localisation of FLAG-tagged BRCA2 protein. Whilst BRCA2 protein is equally distributed between the nucleus and the cytoplasm in 14% of cells, it is predominantly nuclear (A) in 42% and cytoplasmic (B) in 44% of cells (n=50). Expression in the cytoplasm appears granular, which may indicate an association with the endoplasmic reticulum. Staining is not seen in untransfected cells.

Detailed Description

Preparation BRCA2 nucleic acid, and vectors and host cells incorporating the nucleic acid

"BRCA2 region" refers to the portion of human chromosome 13q12-13 identified in Wooster et al (5), containing the BRCA2 locus.

The "BRCA2 locus" includes the BRCA2 gene, both the coding sequence (exons) and intervening sequences (introns), and its regulatory elements for controlling transcription and/or translation. The BRCA2 locus covers allelic variations within the locus.

The term "BRCA2 gene" or "BRCA2 nucleic acid" includes normal alleles of the BRCA2 gene, both silent alleles having no effect on the amino acid sequence of the BRCA2 polypeptide and alleles leading to amino acid sequence variants of BRCA2 polypeptide that do not substantially affect its function. These terms also includes alleles having one or more mutations that are linked to a predisposition to develop cancer, especially male and female breast cancer or ovarian cancer. A mutation may be a change in the BRCA2 nucleic acid sequence which produces a deleterious change in the amino acid sequence of the BRCA2 polypeptide, resulting in partial or complete loss of BRCA2 function, or may be a change in the nucleic acid sequence which results in the loss of effective BRCA2 expression or the production of aberrant forms of the BRCA2 polypeptide. Examples of such mutations are shown in table 1. Alleles including such mutations are also known in the art as susceptibility alleles.

The BRCA2 nucleic acid may be that shown in figures 1, 2, 4 or 7, or it may be an allele as described above, or a variant or derivative, differing from that shown by a change which is one or more of addition, insertion, deletion and substitution of one or more nucleotides of the sequence shown. Changes to a nucleotide sequence may result in an amino acid change at the protein level, or not, as determined by the genetic code.

Thus, nucleic acid according to the present invention may include a sequence different from the sequence shown in figures 1, 2, 4 or 7 yet encode a polypeptide with the same amino acid sequence. The amino acid sequence shown in figure 7 consists of 3418 residues.

On the other hand, the encoded polypeptide may comprise an amino acid sequence which differs by one or more amino acid residues from the amino acid sequence shown in figures 3, 5 or 7. Nucleic acid encoding a polypeptide which is an amino acid sequence variant, derivative or allele of the sequence shown in figures 3, 5 or 7 is further provided by the present invention. Such polypeptides are discussed below. Nucleic acid encoding such a polypeptide may show greater than about 60% homology with the coding sequence shown in figures 1, 2, 4 or 7

greater than about 70% homology, greater than about 80% homology, greater than about 90% homology or greater than about 95% homology.

Particular mutant alleles of the present invention are set out in table 1, using the nomenclature first proposed in (8). These mutations are generally associated with the production of truncated forms of the BRCA2 gene product. These have been shown by the experimental work described herein to be associated with susceptibility to male and female breast cancer and/or ovarian cancer. Implications for screening, e.g. for diagnostic or prognostic purposes, are discussed below.

Generally, nucleic acid according to the present invention is provided as an isolate, in isolated and/or purified form, or free or substantially free of material with which it is naturally associated, such as free or substantially free of nucleic acid flanking the gene in the human genome, except possibly one or more regulatory sequence(s) for expression. Nucleic acid may be wholly or partially synthetic and may include genomic DNA, cDNA or RNA. Where nucleic acid according to the invention includes RNA, reference to the sequence shown should be construed as reference to the RNA equivalent, with U substituted for T.

Nucleic acid sequences encoding all or part of the BRCA2 gene and/or its regulatory elements can be readily prepared by the skilled person using the information and references contained herein and techniques known in the art (for example, see Sambrook, Fritsch and Maniatis, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, and Ausubel et al, Short Protocols in Molecular Biology, John Wiley and Sons, 1992). These techniques include (i) the use of the polymerase chain reaction (PCR) to amplify samples of such nucleic acid, e.g. from genomic sources, (ii) chemical synthesis, or (iii) preparing cDNA sequences. Modifications to the BRCA2 sequences can be made, e.g. using site directed mutagenesis, to provide expression of modified BRCA2 polypeptide or to take account of codon preference in the host cells used to express the nucleic acid.

In order to obtain expression of the BRCA2 nucleic acid sequences, the sequences can be incorporated in a vector having control sequences operably linked to the BRCA2 nucleic acid to control its expression. The vectors may include other sequences such as promoters or enhancers to drive the expression of the inserted nucleic acid, nucleic acid sequences so that the BRCA2 polypeptide is produced as a fusion and/or nucleic acid encoding secretion signals so that the polypeptide produced in the host cell is secreted from the cell. BRCA2 polypeptide can then be obtained by transforming the vectors into host cells in

which the vector is functional, culturing the host cells so that the BRCA2 polypeptide is produced and recovering the BRCA2 polypeptide from the host cells or the surrounding medium. Prokaryotic and eukaryotic cells are used for this purpose in the art, including strains of E. coli, yeast, and eukaryotic cells such as COS or CHO cells. The choice of host cell can be used to control the properties of the BRCA2 polypeptide expressed in those cells, e.g. controlling where the polypeptide is deposited in the host cells or affecting properties such as its glycosylation.

PCR techniques for the amplification of nucleic acid are described in US Patent No. 4,683,195. In general, such techniques require that sequence information from the ends of the target sequence is known to allow suitable forward and reverse oligonucleotide primers to be designed to be identical or similar to the polynucleotide sequence that is the target for the amplification. PCR comprises steps of denaturation of template nucleic acid (if double-stranded), annealing of primer to target, and polymerisation. The nucleic acid probed or used as template in the amplification reaction may be genomic DNA, cDNA or RNA. PCR can be used to amplify specific sequences from genomic DNA, specific RNA sequences and cDNA transcribed from mRNA, bacteriophage or plasmid sequences. The BRCA2 nucleic acid sequences provided herein readily allow the skilled person to design PCR primers, see for example figure 8. References for the general use of PCR techniques include Mullis et al, Cold Spring Harbor Symp. Quant. Biol., 51:263, (1987), Ehrlich (ed), PCR technology, Stockton Press, NY, 1989, Ehrlich et al, Science, 252:1643-1650, (1991), PCR protocols; A Guide to Methods and Applications, Eds. Innis et al, Academic Press, New York, (1990).

Also included within the scope of the invention are antisense oligonucleotide sequences based on the BRCA2 nucleic acid sequences described herein. Antisense oligonucleotides may be designed to hybridise to the complementary sequence of nucleic acid, pre-mRNA or mature mRNA, interfering with the production of polypeptide encoded by a given DNA sequence (e.g. either native BRCA2 polypeptide or a mutant form thereof), so that its expression is reduce or prevented altogether. In addition to the BRCA2 coding sequence, antisense techniques can be used to target the control sequences of the BRCA2 gene, e.g. in the 5' flanking sequence of the BRCA2 coding sequence, whereby the antisense oligonucleotides can interfere with BRCA2 control sequences. The construction of antisense sequences and their use is described in Peyman and Ulman, Chemical Reviews, 90:543-584, (1990), Crooke, Ann. Rev. Pharmacol. Toxicol., 32:329-376, (1992), and Zamecnik and Stephenson, P.N.A.S, 75:280-284, (1974).

The nucleic acid sequences provided in figures 1, 2, 4 and 7 are useful for identifying nucleic acid of interest (and which may be according to the present invention) in a test sample. The present invention provides a method of obtaining nucleic acid of interest, the method including hybridisation of a probe having the sequence shown in figures 1, 2, 4 or 7 or a complementary sequence, to target nucleic acid.

Hybridisation is generally followed by identification of successful hybridisation and isolation of nucleic acid which has hybridised to the probe, which may involve one or more steps of PCR.

Nucleic acid according to the present invention is obtainable using one or more oligonucleotide probes or primers designed to hybridise with one or more fragments of the nucleic acid sequence shown in figures 1, 2, 4 or 7, particularly fragments of relatively rare sequence, based on codon usage or statistical analysis. A primer designed to hybridise with a fragment of the nucleic acid sequence shown in the above figures may be used in conjunction with one or more oligonucleotides designed to hybridise to a sequence in a cloning vector within which target nucleic acid has been cloned, or in so-called "RACE" (rapid amplification of cDNA ends) in which cDNA's in a library are ligated to an oligonucleotide linker and PCR is performed using a primer which hybridises with the sequence shown in figures 1, 2, 4 or 7 and a primer which hybridises to the oligonucleotide linker.

Such oligonucleotide probes or primers, as well as the full-length sequence (and alleles, variants and derivatives) are also useful in screening a test sample containing nucleic acid for the presence of alleles and variants, especially those that confer susceptibility or predisposition to cancers, the probes hybridising with a target sequence from a sample obtained from the individual being tested. The conditions of the hybridisation can be controlled to minimise non-specific binding, and preferably stringent to moderately stringent hybridisation conditions are preferred. The skilled person is readily able to design such probes, label them and devise suitable conditions for the hybridisation reactions, assisted by textbooks such as Sambrook et al (1989) and Ausubel et al (1992).

As well as determining the presence of polymorphisms or mutations in the BRCA2 sequence, the probes may also be used to determine whether mRNA encoding BRCA2 is present in a cell or tissue.

Nucleic acid isolated and/or purified from one or more cells (e.g. human) or a nucleic acid library derived from nucleic acid isolated and/or purified from cells (e.g. a cDNA library derived from mRNA

isolated from the cells), may be probed under conditions for selective hybridisation and/or subjected to a specific nucleic acid amplification reaction such as the polymerase chain reaction (PCR).

In the context of cloning, it may be necessary for one or more gene fragments to be ligated to generate a full-length coding sequence. Also, where a full-length encoding nucleic acid molecule has not been obtained, a smaller molecule representing part of the full molecule, may be used to obtain full-length clones. Inserts may be prepared from partial cDNA clones and used to screen cDNA libraries. The full-length clones isolated may be subcloned into expression vectors and activity assayed by transfection into suitable host cells, e.g. with a reporter plasmid.

A method may include hybridisation of one or more (e.g. two) probes or primers to target nucleic acid. Where the nucleic acid is double-stranded DNA, hybridisation will generally be preceded by denaturation to produce single-stranded DNA. The hybridisation may be as part of a PCR procedure, or as part of a probing procedure not involving PCR. An example procedure would be a combination of PCR and low stringency hybridisation. A screening procedure, chosen from the many available to those skilled in the art, is used to identify successful hybridisation events and isolated hybridised nucleic acid.

Binding of a probe to target nucleic acid (e.g. DNA) may be measured using any of a variety of techniques at the disposal of those skilled in the art. For instance, probes may be radioactively, fluorescently or enzymatically labelled. Other methods not employing labelling of probe include examination of restriction fragment length polymorphisms, amplification using PCR, RNAase cleavage and allele specific oligonucleotide probing.

Probing may employ the standard Southern blotting technique. For instance DNA may be extracted from cells and digested with different restriction enzymes. Restriction fragments may then be separated by electrophoresis on an agarose gel, before denaturation and transfer to a nitrocellulose filter. Labelled probe may be hybridised to the DNA fragments on the filter and binding determined. DNA for probing may be prepared from RNA preparations from cells.

Preliminary experiments may be performed by hybridising under low stringency conditions various probes to Southern blots of DNA digested with restriction enzymes. Suitable conditions would be achieved when a large number of hybridising fragments were obtained while the background hybridisation was low. Using these conditions nucleic acid

libraries, e.g. cDNA libraries representative of expressed sequences, may be searched.

Those skilled in the art are well able to employ suitable conditions of the desired stringency for selective hybridisation, taking into account factors such as oligonucleotide length and base composition, temperature and so on.

On the basis of amino acid sequence information, oligonucleotide probes or primers may be designed, taking into account the degeneracy of the genetic code, and, where appropriate, codon usage of the organism from the candidate nucleic acid is derived. An oligonucleotide for use in nucleic acid amplification may have about 10 or fewer codons (e.g. 6, 7 or 8), i.e. be about 30 or fewer nucleotides in length (e.g. 18, 21 or 24). Generally specific primers are upwards of 14 nucleotides in length, but not more than 18-20. Those skilled in the art are well versed in the design of primers for use processes such as PCR.

A further aspect of the present invention provides an oligonucleotide or polynucleotide fragment of the nucleotide sequence shown in figure 1, 2, 4 or 7, or a complementary sequence, in particular for use in a method of obtaining and/or screening nucleic acid. The sequences referred to above may be modified by addition, substitution, insertion or deletion of one or more nucleotides, but preferably without abolition of ability to hybridise selectively with nucleic acid with the sequence shown in figures 1, 2, 4 or 7, that is wherein the degree of homology of the oligonucleotide or polynucleotide with one of the sequences given is sufficiently high.

In some preferred embodiments, oligonucleotides according to the present invention that are fragments of any of the sequences shown in figures 1, 2, 4 or 7, or any allele associated with cancer susceptibility, are at least about 10 nucleotides in length, more preferably at least about 15 nucleotides in length, more preferably at least about 20 nucleotides in length. Such fragments themselves individually represent aspects of the present invention. Fragments and other oligonucleotides may be used as primers or probes as discussed but may also be generated (e.g. by PCR) in methods concerned with determining the presence in a test sample of a sequence indicative of cancer susceptibility.

Methods involving use of nucleic acid in diagnostic and/or prognostic contexts, for instance in determining susceptibility to cancer, and other methods concerned with determining the presence of sequences indicative of cancer susceptibility are discussed below.

Nucleic acid according to the present invention may be used in methods of gene therapy, for instance in treatment of individuals with the aim of preventing or curing (wholly or partially) cancer. This too is discussed below.

Nucleic acid according to the present invention, such as a full-length coding sequence or oligonucleotide probe or primer, may be provided as part of a kit, e.g. in a suitable container such as a vial in which the contents are protected from the external environment. The kit may include instructions for use of the nucleic acid, e.g. in PCR and/or a method for determining the presence of nucleic acid of interest in a test sample. A kit wherein the nucleic acid is intended for use in PCR may include one or more other reagents required for the reaction, such as polymerase, nucleosides, buffer solution etc. The nucleic acid may be labelled. A kit for use in determining the presence or absence of nucleic acid of interest may include one or more articles and/or reagents for performance of the method, such as means for providing the test sample itself, e.g. a swab for removing cells from the buccal cavity or a syringe for removing a blood sample (such components generally being sterile). In a further aspect, the present invention provides an apparatus for screening BRCA2 nucleic acid, the apparatus comprising storage means including the BRCA2 nucleic acid sequence as set out in any one of figures 1, 2, 4 or 7, the stored sequence being used to compare the sequence of the test nucleic acid to determine the presence of mutations.

A convenient way of producing a polypeptide according to the present invention is to express nucleic acid encoding it, by use of the nucleic acid in an expression system. The use of expression system has reached an advanced degree of sophistication today.

Accordingly, the present invention also encompasses a method of making a polypeptide (as disclosed), the method including expression from nucleic acid encoding the polypeptide (generally nucleic acid according to the invention). This may conveniently be achieved by growing a host cell in culture, containing such a vector, under appropriate conditions which cause or allow expression of the polypeptide. Polypeptides may also be expressed in in vitro systems, such as reticulocyte lysate.

Systems for cloning and expression of a polypeptide in a variety of different host cells are well known. Suitable host cells include bacteria, eukaryotic cells such as mammalian and yeast, and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells, COS cells and many others. A common,

preferred bacterial host is E. coli.

Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmids, viral e.g. 'phage, or phagemid, as appropriate. For further details see, for example, Molecular Cloning: a Laboratory Manual: 2nd edition, Sambrook et al., 1989, Cold Spring Harbor Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in Current Protocols in Molecular Biology, Ausubel et al. eds., John Wiley & Sons, 1992.

Thus, a further aspect of the present invention provides a host cell containing nucleic acid as disclosed herein. The nucleic acid of the invention may be integrated into the genome (e.g. chromosome) of the host cell. Integration may be promoted by inclusion of sequences which promote recombination with the genome, in accordance with standard techniques. The nucleic acid may be on an extra-chromosomal vector within the cell.

A still further aspect provides a method which includes introducing the nucleic acid into a host cell. The introduction, which may (particularly for in vitro introduction) be generally referred to without limitation as "transformation", may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transfection using bacteriophage. As an alternative, direct injection of the nucleic acid could be employed.

Marker genes such as antibiotic resistance or sensitivity genes may be used in identifying clones containing nucleic acid of interest, as is well known in the art.

The introduction may be followed by causing or allowing expression from the nucleic acid, e.g. by culturing host cells (which may include cells actually transformed although more likely the cells will be descendants of the transformed cells) under conditions for expression of the gene, so that the encoded polypeptide is produced. If the polypeptide is

expressed coupled to an appropriate signal leader peptide it may be secreted from the cell into the culture medium. Following production by expression, a polypeptide may be isolated and/or purified from the host cell and/or culture medium, as the case may be, and subsequently used as desired, e.g. in the formulation of a composition which may include one or more additional components, such as a pharmaceutical composition which includes one or more pharmaceutically acceptable excipients, vehicles or carriers (e.g. see below).

Introduction of nucleic acid may take place in vivo by way of gene therapy, as discussed below.

A host cell containing nucleic acid according to the present invention, e.g. as a result of introduction of the nucleic acid into the cell or into an ancestor of the cell and/or genetic alteration of the sequence endogenous to the cell or ancestor (which introduction or alteration may take place in vivo or ex vivo), may be comprised (e.g. in the soma) within an organism which is an animal, particularly a mammal, which may be human or non-human, such as rabbit, guinea pig, rat, mouse or other rodent, cat, dog, pig, sheep, goat, cattle or horse, or which is a bird, such as a chicken. Genetically modified or transgenic animals or birds comprising such a cell are also provided as further aspects of the present invention.

This may have a therapeutic aim. (Gene therapy is discussed below.) The presence of a mutant, allele or variant sequence within cells of an organism, particularly when in place of a homologous endogenous sequence, may allow the organism to be used as a model in testing and/or studying the role of the BRCA2 gene or substances which modulate activity of the encoded polypeptide in vitro or the promoter sequence shown in figure 6 are otherwise indicated to be of therapeutic potential.

Instead of or as well as being used for the production of a polypeptide encoded by a transgene, host cells may be used as a nucleic acid factory to replicate the nucleic acid of interest in order to generate large amounts of it. Multiple copies of nucleic acid of interest may be made within a cell when coupled to an amplifiable gene such as DHFR. Host cells transformed with nucleic acid of interest, or which are descended from host cells into which nucleic acid was introduced, may be cultured under suitable conditions, e.g. in a fermenter, taken from the culture and subjected to processing to purify the nucleic acid. Following purification, the nucleic acid or one or more fragments thereof may be used as desired, for instance in a diagnostic or prognostic assay as discussed elsewhere herein.

Production of BRCA2 Polypeptides

The skilled person can use the techniques described herein and others well known in the art to produce large amounts of the BRCA2 polypeptide, or fragments or active portions thereof, for use as pharmaceuticals, in the developments of drugs and for further study into its properties and role in vivo. Experimental work confirming the production of BRCA2 polypeptide is set out in example 3 below.

Thus, a further aspect of the present invention provides a polypeptide which has the amino acid sequence shown in figures 3, 5 or 7, which may be in isolated and/or purified form, free or substantially free of material with which it is naturally associated, such as other polypeptides or such as human polypeptides other than BRCA2 polypeptide or (for example if produced by expression in a prokaryotic cell) lacking in native glycosylation, e.g. unglycosylated.

Polypeptides which are amino acid sequence variants, alleles or derivatives are also provided by the present invention. A polypeptide which is a variant, allele, or derivative may have an amino acid sequence which differs from that given in figures 3, 5 or 7 by one or more of addition, substitution, deletion and insertion of one or more amino acids. Preferred such polypeptides have BRCA2 function, that is to say have one or more of the following properties: immunological cross-reactivity with an antibody reactive the polypeptide for which the sequence is given in figures 3, 5 or 7; sharing an epitope with the polypeptide for which the amino acid sequence is shown in figures 3, 5 or 7 (as determined for example by immunological cross-reactivity between the two polypeptides.

A polypeptide which is an amino acid sequence variant, or allele, derivative of the amino acid sequence shown in figures 3, 5, or 7 may comprise an amino acid sequence which shares greater than about 35% sequence identity with the sequence shown, greater than about 40%, greater than about 50%, greater than about 60%, greater than about 70%, greater than about 80%, greater than about 90% or greater than about 95%. The sequence may share greater than about 60% similarity, greater than about 70% similarity, greater than about 80% similarity or greater than about 90% similarity with the amino acid sequence shown in any one of figures 3, 5 or 7. Particular amino acid sequence variants may differ from those shown in figures 3, 5 or 7 by insertion, addition, substitution or deletion of 1 amino acid, 2, 3, 4, 5-10, 10-20 20-30, 30-50, 50-100, 100-150, or more than 150 amino acids.

By way of example, mutation variants representative of preferred

embodiments of the present invention are shown in table 1. Screening for the presence of one or more of these in a test sample has a diagnostic and/or prognostic use, for instance in determining cancer susceptibility, as discussed below.

The present invention also includes active portions, fragments, derivatives and functional mimetics of the BRCA2 polypeptides of the invention.

An "active portion" of BRCA2 polypeptide means a peptide which is less than said full length BRCA2 polypeptide, but which retains its essential biological activity.

A "fragment" of the BRCA2 polypeptide means a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids. Fragments of the BRCA2 polypeptide sequence antigenic determinants or epitopes useful for raising antibodies to a portion of the BRCA2 amino acid sequence.

A "derivative" of the BRCA2 polypeptide or a fragment thereof means a polypeptide modified by varying the amino acid sequence of the protein, e.g. by manipulation of the nucleic acid encoding the protein or by altering the protein itself. Such derivatives of the natural amino acid sequence may involve insertion, addition, deletion or substitution of one or more amino acids, without fundamentally altering the essential activity of the wild type BRCA2 polypeptide.

"Functional mimetic" means a substance which may not contain an active portion of the BRCA2 amino acid sequence, and probably is not a peptide at all, but which retains the essential biological activity of natural BRCA2 polypeptide. The design and screening of candidate mimetics is described in detail below.

A polypeptide according to the present invention may be isolated and/or purified (e.g. using an antibody) for instance after production by expression from encoding nucleic acid (for which see below). Polypeptides according to the present invention may also be generated wholly or partly by chemical synthesis. The isolated and/or purified polypeptide may be used in formulation of a composition, which may include at least one additional component, for example a pharmaceutical composition including a pharmaceutically acceptable excipient, vehicle or carrier. A composition including a polypeptide according to the invention may be used in prophylactic and/or therapeutic treatment as

discussed below.

A polypeptide, peptide fragment, allele, or variant according to the present invention may be used as an immunogen or otherwise in obtaining specific antibodies. Antibodies are useful in purification and other manipulation of polypeptides and peptides, diagnostic screening and therapeutic contexts. This is discussed further below.

A polypeptide according to the present invention may be used in screening for molecules which affect or modulate its activity or function. Such molecules may be useful in a therapeutic (possibly including prophylactic) context.

Production of BRCA2 Antibodies

A further important use of the BRCA2 polypeptides is in raising antibodies that have the property of specifically binding to the BRCA2 polypeptides, or fragments or active portions thereof.

The production of monoclonal antibodies is well established in the art. Monoclonal antibodies can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP-A-184187, GB-A-2188638 or EP-A-239400. A hybridoma producing a monoclonal antibody may be subject to genetic mutation or other changes, which may or may not alter the binding specificity of antibodies produced.

The provision of the novel BRCA2 polypeptides enables for the first time the production of antibodies able to bind it specifically. Accordingly, a further aspect of the present invention provides an antibody able to bind specifically to the polypeptide whose sequence is given in figures 3, 5 or 7. Such an antibody may be specific in the sense of being able to distinguish between the polypeptide it is able to bind and other human polypeptides for which it has no or substantially no binding affinity (e.g. a binding affinity of about 1000x worse). Specific antibodies bind an epitope on the molecule which is either not present or is not accessible on other molecules. Antibodies according to the present invention may be specific for the wild-type polypeptide. Antibodies according to the invention may be specific for a particular variant, allele or derivative polypeptide as between that molecule and the wild-type BRCA2 polypeptide, so as to be

useful in diagnostic and prognostic methods as discussed below. Antibodies are also useful in purifying the polypeptide or polypeptides to which they bind, e.g. following production by recombinant expression from encoding nucleic acid.

Preferred antibodies according to the invention are isolated, in the sense of being free from contaminants such as antibodies able to bind other polypeptides and/or free of serum components. Monoclonal antibodies are preferred for some purposes, though polyclonal antibodies are within the scope of the present invention.

Antibodies may be obtained using techniques which are standard in the art. Methods of producing antibodies include immunising a mammal (e.g. mouse, rat, rabbit, horse, goat, sheep or monkey) with the protein or a fragment thereof. Antibodies may be obtained from immunised animals using any of a variety of techniques known in the art, and screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or immunoprecipitation may be used (Armitage et al, Nature, 357:80-82, 1992). Isolation of antibodies and/or antibody-producing cells from an animal may be accompanied by a step of sacrificing the animal.

As an alternative or supplement to immunising a mammal with a peptide, an antibody specific for a protein may be obtained from a recombinantly produced library of expressed immunoglobulin variable domains, e.g. using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047. The library may be naive, that is constructed from sequences obtained from an organism which has not been immunised with any of the proteins (or fragments), or may be one constructed using sequences obtained from an organism which has been exposed to the antigen of interest.

Antibodies according to the present invention may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any binding substance having a binding domain with the required specificity. Thus the invention covers antibody fragments, derivatives, functional equivalents and homologues of antibodies, including synthetic molecules and molecules whose shape mimics that of an antibody enabling it to bind an antigen or epitope.

Example antibody fragments, capable of binding an antigen or other binding partner are the Fab fragment consisting of the VL, VH, Cl and CH1 domains; the Fd fragment consisting of the VH and CH1 domains; the Fv fragment consisting of the VL and VH domains of a single arm of an

antibody; the dAb fragment which consists of a VH domain; isolated CDR regions and F(ab')2 fragments, a bivalent fragment including two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included.

Humanised antibodies in which CDRs from a non-human source are grafted onto human framework regions, typically with the alteration of some of the framework amino acid residues, to provide antibodies which are less immunogenic than the parent non-human antibodies, are also included within the present invention.

A hybridoma producing a monoclonal antibody according to the present invention may be subject to genetic mutation or other changes. It will further be understood by those skilled in the art that a monoclonal antibody can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP-A-184187, GB-A-2188638 or EP-A-0239400. Cloning and expression of chimeric antibodies are described in EP-A-0120694 and EP-A-0125023.

Hybridomas capable of producing antibody with desired binding characteristics are within the scope of the present invention, as are host cells, eukaryotic or prokaryotic, containing nucleic acid encoding antibodies (including antibody fragments) and capable of their expression. The invention also provides methods of production of the antibodies including growing a cell capable of producing the antibody under conditions in which the antibody is produced, and preferably secreted.

The reactivities of antibodies on a sample may be determined by any appropriate means. Tagging with individual reporter molecules is one possibility. The reporter molecules may directly or indirectly generate detectable, and preferably measurable, signals. The linkage of reporter molecules may be directly or indirectly, covalently, e.g. via a peptide bond or non-covalently. Linkage via a peptide bond may be as a result of recombinant expression of a gene fusion encoding antibody and reporter molecule.

One favoured mode is by covalent linkage of each antibody with an individual fluorochrome, phosphor or laser dye with spectrally isolated absorption or emission characteristics. Suitable fluorochromes include

fluorescein, rhodamine, phycoerythrin and Texas Red. Suitable chromogenic dyes include diaminobenzidine.

Other reporters include macromolecular colloidal particles or particulate material such as latex beads that are coloured, magnetic or paramagnetic, and biologically or chemically active agents that can directly or indirectly cause detectable signals to be visually observed, electronically detected or otherwise recorded. These molecules may be enzymes which catalyse reactions that develop or change colours or cause changes in electrical properties, for example. They may be molecularly excitable, such that electronic transitions between energy states result in characteristic spectral absorptions or emissions. They may include chemical entities used in conjunction with biosensors. Biotin/avidin or biotin/streptavidin and alkaline phosphatase detection systems may be employed.

The mode of determining binding is not a feature of the present invention and those skilled in the art are able to choose a suitable mode according to their preference and general knowledge.

Antibodies according to the present invention may be used in screening for the presence of a polypeptide, for example in a test sample containing cells or cell lysate as discussed, and may be used in purifying and/or isolating a polypeptide according to the present invention, for instance following production of the polypeptide by expression from encoding nucleic acid therefor. Antibodies may modulate the activity of the polypeptide to which they bind and so, if that polypeptide has a deleterious effect in an individual, may be useful in a therapeutic context (which may include prophylaxis).

An antibody may be provided in a kit, which may include instructions for use of the antibody, e.g. in determining the presence of a particular substance in a test sample. One or more other reagents may be included, such as labelling molecules, buffer solutions, elutants and so on. Reagents may be provided within containers which protect them from the external environment, such as a sealed vial.

Diagnostic Methods

A number of methods are known in the art for analysing biological samples from individuals to determine whether the individual carries a BRCA2 allele predisposing them to cancer, especially breast cancer (female or male) or ovarian cancer. The purpose of such analysis may be used for diagnosis or prognosis, and serve to detect the presence of an existing cancer, to help identify the type of cancer, to assist a

physician in determining the severity or likely course of the cancer and/or to optimise treatment of it. Alternatively, the methods can be used to detect BRCA2 alleles that are statistically associated with a susceptibility to cancer in the future, e.g. early onset breast cancer, identifying individuals who would benefit from regular screening to provide early diagnosis of cancer. Examples of methods of screening for BRCA2 mutations are set out in example 5 below.

Broadly, the methods divide into those screening for the presence of BRCA2 nucleic acid sequences or alleles or variants thereof, and those that rely on detecting the presence or absence of the BRCA2 polypeptide. The methods make use of biological samples from individuals that are suspected of contain the nucleic acid sequences or polypeptide. Examples of biological samples include blood, plasma, serum, tissue samples, tumour samples, saliva and urine.

Exemplary approaches for detecting BRCA2 nucleic acid or polypeptides include:

(a) comparing the sequence of nucleic acid in the sample with the BRCA2 nucleic acid sequence to determine whether the sample from the patient contains mutations; or,

(b) determining the presence in a sample from a patient of the polypeptide encoded by the BRCA2 gene and, if present, determining whether the polypeptide is full length, and/or is mutated, and/or is expressed at the normal level; or,

(c) using DNA fingerprinting to compare the restriction pattern produced when a restriction enzyme cuts a sample of nucleic acid from the patient with the restriction pattern obtained from normal BRCA2 gene or from known mutations thereof; or,

(d) using a specific binding member capable of binding to a BRCA2 nucleic acid sequence (either a normal sequence or a known mutated sequence), the specific binding member comprising nucleic acid hybridisable with the BRCA2 sequence, or substances comprising an antibody domain with specificity for a native or mutated BRCA2 nucleic acid sequence or the polypeptide encoded by it, the specific binding member being labelled so that binding of the specific binding member to its binding partner is detectable; or,

(e) using PCR involving one or more primers based on normal or mutated BRCA2 gene sequence to screen for normal or mutant BRCA2 gene in a sample from a patient.

A "specific binding pair" comprises a specific binding member (sbm) and a binding partner (bp) which have a particular specificity for each other and which in normal conditions bind to each other in preference to other molecules. Examples of specific binding pairs are antigens

and antibodies, molecules and receptors and complementary nucleotide sequences. The skilled person will be able to think of many other examples and they do not need to be listed here. Further, the term "specific binding pair" is also applicable where either or both of the specific binding member and the binding partner comprise a part of a larger molecule. In embodiments in which the specific binding pair are nucleic acid sequences, they will be of a length to hybridise to each other under the conditions of the assay, preferably greater than 10 nucleotides long, more preferably greater than 15 or 20 nucleotides long.

In most embodiments for screening for BRCA2 susceptibility alleles, the BRCA2 nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other sequences present in the sample. This allows the target sequences to be detected with a high degree of sensitivity if they are present in the sample. This initial step may be avoided by using highly sensitive array techniques that are becoming increasingly important in the art.

The identification of the BRCA2 gene and its association with cancer paves the way for aspects of the present invention to provide the use of materials and methods, such as are disclosed and discussed above, for establishing the presence or absence in a test sample of an variant form of the gene, in particular an allele or variant specifically associated with cancer, especially breast and ovarian cancer. This may be for diagnosing a predisposition of an individual to cancer. It may be for diagnosing cancer of a patient with the disease as being associated with the gene.

This allows for planning of appropriate therapeutic and/or prophylactic treatment, permitting stream-lining of treatment by targeting those most likely to benefit.

A variant form of the gene may contain one or more insertions, deletions, substitutions and/or additions of one or more nucleotides compared with the wild-type sequence (such as shown in table 1) which may or may not disrupt the gene function. Differences at the nucleic acid level are not necessarily reflected by a difference in the amino acid sequence of the encoded polypeptide. However, a mutation or other difference in a gene may result in a frame-shift or stop codon, which could seriously affect the nature of the polypeptide produced (if any), or a point mutation or gross mutational change to the encoded polypeptide, including insertion, deletion, substitution and/or addition of one or more amino acids or regions in the polypeptide. A mutation in a promoter sequence or other regulatory region may prevent

or reduce expression from the gene or affect the processing or stability of the mRNA transcript.

There are various methods for determining the presence or absence in a test sample of a particular nucleic acid sequence, such as the sequence shown in figures 1, 2, 4 or 7 or a variant or allele thereof.

Tests may be carried out on preparations containing genomic DNA, cDNA and/or mRNA. Testing cDNA or mRNA has the advantage of the complexity of the nucleic acid being reduced by the absence of intron sequences, but the possible disadvantage of extra time and effort being required in making the preparations. RNA is more difficult to manipulate than DNA because of the wide-spread occurrence of RN'ases.

Nucleic acid in a test sample may be sequenced and the sequence compared with the sequence shown in figures 1, 2, 4 or 7, to determine whether or not a difference is present. If so, the difference can be compared with known susceptibility alleles (e.g. as summarised in table 1) to determine whether the test nucleic acid contains one or more of the variations indicated, or the difference can be investigated for association with cancer.

Since it will not generally be time- or labour-efficient to sequence all nucleic acid in a test sample or even the whole BRCA2 gene, a specific amplification reaction such as PCR using one or more pairs of primers may be employed to amplify the region of interest in the nucleic acid, for instance the BRCA2 gene or a particular region in which mutations associated with cancer susceptibility occur. Exemplary primers for this purpose are shown in figure 8. The amplified nucleic acid may then be sequenced as above, and/or tested in any other way to determine the presence or absence of a particular feature. Nucleic acid for testing may be prepared from nucleic acid removed from cells or in a library using a variety of other techniques such as restriction enzyme digest and electrophoresis.

Nucleic acid may be screened using a variant- or allele-specific probe. Such a probe corresponds in sequence to a region of the BRCA2 gene, or its complement, containing a sequence alteration known to be associated with cancer susceptibility. Under suitably stringent conditions, specific hybridisation of such a probe to test nucleic acid is indicative of the presence of the sequence alteration in the test nucleic acid. For efficient screening purposes, more than one probe may be used on the same test sample.

Allele- or variant-specific oligonucleotides may similarly be used in

PCR to specifically amplify particular sequences if present in a test sample. Assessment of whether a PCR band contains a gene variant may be carried out in a number of ways familiar to those skilled in the art. The PCR product may for instance be treated in a way that enables one to display the mutation or polymorphism on a denaturing polyacrylamide DNA sequencing gel, with specific bands that are linked to the gene variants being selected.

An alternative or supplement to looking for the presence of variant sequences in a test sample is to look for the presence of the normal sequence, e.g. using a suitably specific oligonucleotide probe or primer.

Use of oligonucleotide probes and primers has been discussed in more detail above.

Approaches which rely on hybridisation between a probe and test nucleic acid and subsequent detection of a mismatch may be employed. Under appropriate conditions (temperature, pH etc.), an oligonucleotide probe will hybridise with a sequence which is not entirely complementary. The degree of base-pairing between the two molecules will be sufficient for them to anneal despite a mis-match. Various approaches are well known in the art for detecting the presence of a mis-match between two annealing nucleic acid molecules.

For instance, RN'ase A cleaves at the site of a mis-match. Cleavage can be detected by electrophoresing test nucleic acid to which the relevant probe or probe has annealed and looking for smaller molecules (i.e. molecules with higher electrophoretic mobility) than the full length probe/test hybrid. Other approaches rely on the use of enzymes such as resolvases or endonucleases.

Thus, an oligonucleotide probe that has the sequence of a region of the normal BRCA2 gene (either sense or anti-sense strand) in which mutations associated with cancer susceptibility are known to occur (e.g. see table 1) may be annealed to test nucleic acid and the presence or absence of a mis-match determined. Detection of the presence of a mis-match may indicate the presence in the test nucleic acid of a mutation associated with cancer susceptibility. On the other hand, an oligonucleotide probe that has the sequence of a region of the BRCA2 gene including a mutation associated with cancer susceptibility may be annealed to test nucleic acid and the presence or absence of a mis-match determined. The absence of a mis-match may indicate that the nucleic acid in the test sample has the normal sequence. In either case, a battery of probes to different regions of the gene may be

employed.

The presence of differences in sequence of nucleic acid molecules may be detected by means of restriction enzyme digestion, such as in a method of DNA fingerprinting where the restriction pattern produced when one or more restriction enzymes are used to cut a sample of nucleic acid is compared with the pattern obtained when a sample containing the normal gene or a variant or allele is digested with the same enzyme or enzymes.

The presence of absence of a lesion in a promoter or other regulatory sequence may also be assessed by determining the level of mRNA production by transcription or the level of polypeptide production by translation from the mRNA.

A test sample of nucleic acid may be provided for example by extracting nucleic acid from cells, e.g. in saliva or preferably blood, or for pre-natal testing from the amnion, placenta or foetus itself.

There are various methods for determining the presence or absence in a test sample of a particular polypeptide, such as the polypeptide with the amino acid sequence shown in figure 3, 5 or 7 or an amino acid sequence variant or allele thereof.

A sample may be tested for the presence of a binding partner for a specific binding member such as an antibody (or mixture of antibodies), specific for one or more particular variants of the polypeptide shown in figures 3, 5 or 7.

A sample may be tested for the presence of a binding partner for a specific binding member such as an antibody (or mixture of antibodies), specific for the polypeptide shown in figures 3, 5 or 7.

In such cases, the sample may be tested by being contacted with a specific binding member such as an antibody under appropriate conditions for specific binding, before binding is determined, for instance using a reporter system as discussed. Where a panel of antibodies is used, different reporting labels may be employed for each antibody so that binding of each can be determined.

A specific binding member such as an antibody may be used to isolate and/or purify its binding partner polypeptide from a test sample, to allow for sequence and/or biochemical analysis of the polypeptide to determine whether it has the sequence and/or properties of the polypeptide whose sequence is shown in figures 3, 5 or 7, or if it is

a mutant or variant form. Amino acid sequence is routine in the art using automated sequencing machines.

There is also an increasing tendency in the diagnostic field towards miniaturisation of such assays, e.g. making use of binding agents (such as antibodies or nucleic acid sequences) immobilised in small, discrete locations (microspots) and/or as arrays on solid supports or on diagnostic chips. These approaches can be particularly valuable as they can provide great sensitivity (particularly through the use of fluorescently labelled reagents), require only very small amounts of biological sample from individuals being tested and allow a variety of separate assays can be carried out simultaneously. This latter advantage can be useful as it provides an assay for different mutations in the BRCA2 gene or another cancer susceptibility gene (such as BRCA1, see EP-A-705902) to be carried out using a single sample. Examples of techniques enabling this miniaturised technology are provided in WO84/01031, WO88/1058, WO89/01157, WO93/8472, WO95/18376/ WO95/18377, WO95/24649 and EP-A-0373203. Thus, in a further aspect, the present invention provides a kit comprising a support or diagnostic chip having immobilised thereon one or more binding agents capable of specifically binding BRCA2 nucleic acid or polypeptides, optionally in combination with other reagents (such as labelled developing reagents) needed to carrying out an assay.

Therapeutics

Pharmaceuticals and Peptide Therapies

The BRCA2 polypeptides, antibodies, peptides and nucleic acid of the invention can be formulated in pharmaceutical compositions. These compositions may comprise, in addition to one of the above substances, a pharmaceutically acceptable excipient, carrier, buffer, stabiliser or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, intramuscular, intraperitoneal routes.

Pharmaceutical compositions for oral administration may be in tablet, capsule, powder or liquid form. A tablet may include a solid carrier such as gelatin or an adjuvant. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, dextrose or other saccharide solution or glycols such as

ethylene glycol, propylene glycol or polyethylene glycol may be included.

For intravenous, cutaneous or subcutaneous injection, or injection at the site of affliction, the active ingredient will be in the form of a parenterally acceptable aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Those of relevant skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as sodium chloride injection, Ringer's injection, lactated Ringer's injection. Preservatives, stabilisers, buffers, antioxidants and/or other additives may be included, as required.

Whether it is a polypeptide, antibody, peptide, nucleic acid molecule, small molecule or other pharmaceutically useful compound according to the present invention that is to be given to an individual, administration is preferably in a "prophylactically effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual. The actual amount administered, and rate and time-course of administration, will depend on the nature and severity of what is being treated. Prescription of treatment, e.g. decisions on dosage etc, is within the responsibility of general practitioners and other medical doctors, and typically takes account of the disorder to be treated, the condition of the individual patient, the site of delivery, the method of administration and other factors known to practitioners. Examples of the techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 16th edition, Osol, A. (ed), 1980.

Alternatively, targeting therapies may be used to deliver the active agent more specifically to certain types of cell, by the use of targeting systems such as antibody or cell specific ligands. Targeting may be desirable for a variety of reasons; for example if the agent is unacceptably toxic, or if it would otherwise require too high a dosage, or if it would not otherwise be able to enter the target cells.

Instead of administering these agents directly, they could be produced in the target cells by expression from an encoding gene introduced into the cells, eg in a viral vector (a variant of the VDEPT technique - see below). The vector could be targeted to the specific cells to be treated, or it could contain regulatory elements which are switched on more or less selectively by the target cells.

Alternatively, the agent could be administered in a precursor form, for

conversion to the active form by an activating agent produced in, or targeted to, the cells to be treated. This type of approach is sometimes known as ADEPT or VDEPT; the former involving targeting the activating agent to the cells by conjugation to a cell-specific antibody, while the latter involves producing the activating agent, eg an enzyme, in a vector by expression from encoding DNA in a viral vector (see for example, EP-A-415731 and WO90/07936).

A composition may be administered alone or in combination with other treatments, either simultaneously or sequentially dependent upon the condition to be treated.

Methods of Gene Therapy

As a further alternative, the nucleic acid encoded the authentic biologically active BRCA2 polypeptide could be used in a method of gene therapy, to treat a patient who is unable to synthesize the active polypeptide or unable to synthesize it at the normal level, thereby providing the effect provided by wild-type BRCA2 and suppressing the occurrence of cancer and/or reduce the size or extent of existing cancer in the target cells.

Vectors such as viral vectors have been used in the prior art to introduce genes into a wide variety of different target cells. Typically the vectors are exposed to the target cells so that transfection can take place in a sufficient proportion of the cells to provide a useful therapeutic or prophylactic effect from the expression of the desired polypeptide. The transfected nucleic acid may be permanently incorporated into the genome of each of the targeted tumour cells, providing long lasting effect, or alternatively the treatment may have to be repeated periodically.

A variety of vectors, both viral vectors and plasmid vectors, are known in the art, see US Patent No. 5,252,479 and WO93/07282. In particular, a number of viruses have been used as gene transfer vectors, including papovaviruses, such as SV40, vaccinia virus, herpesviruses, including HSV and EBV, and retroviruses. Many gene therapy protocols in the prior art have used disabled murine retroviruses.

As an alternative to the use of viral vectors other known methods of introducing nucleic acid into cells includes electroporation, calcium phosphate co-precipitation, mechanical techniques such as microinjection, transfer mediated by liposomes and direct DNA uptake and receptor-mediated DNA transfer.

As mentioned above, the aim of gene therapy using nucleic acid encoding the BRCA2 polypeptide, or an active portion thereof, is to increase the amount of the expression product of the nucleic acid in cells in which the level of the wild-type BRCA2 polypeptide is absent or present only at reduced levels. Such treatment may be therapeutic in the treatment of cells which are already cancerous or prophylactic in the treatment of individuals known through screening to have a BRCA2 susceptibility allele and hence a predisposition to cancer.

Gene transfer techniques which selectively target the BRCA2 nucleic acid to breast and/or ovarian tissues are preferred. Examples of this included receptor-mediated gene transfer, in which the nucleic acid is linked to a protein ligand via polylysine, with the ligand being specific for a receptor present on the surface of the target cells.

Antisense technology based on the BRCA2 nucleic acid sequences is discussed above.

Methods of Screening for Drugs

A polypeptide according to the present invention may be used in screening for molecules which affect or modulate its activity or function. Such molecules may be useful in a therapeutic (possibly including prophylactic) context.

It is well known that pharmaceutical research leading to the identification of a new drug may involve the screening of very large numbers of candidate substances, both before and even after a lead compound has been found. This is one factor which makes pharmaceutical research very expensive and time-consuming. Means for assisting in the screening process can have considerable commercial importance and utility. Such means for screening for substances potentially useful in treating or preventing cancer is provided by polypeptides according to the present invention. Substances identified as modulators of the polypeptide represent an advance in the fight against cancer since they provide basis for design and investigation of therapeutics for in vivo use.

A method of screening for a substance which modulates activity of a polypeptide may include contacting one or more test substances with the polypeptide in a suitable reaction medium, testing the activity of the treated polypeptide and comparing that activity with the activity of the polypeptide in comparable reaction medium untreated with the test substance or substances. A difference in activity between the treated and untreated polypeptides is indicative of a modulating effect of the

relevant test substance or substances.

Combinatorial library technology provides an efficient way of testing a potentially vast number of different substances for ability to modulate activity of a polypeptide. Such libraries and their use are known in the art. The use of peptide libraries is preferred.

Prior to or as well as being screened for modulation of activity, test substances may be screened for ability to interact with the polypeptide, e.g. in a yeast two-hybrid system (which requires that both the polypeptide and the test substance can be expressed in yeast from encoding nucleic acid). This may be used as a coarse screen prior to testing a substance for actual ability to modulate activity of the polypeptide. Alternatively, the screen could be used to screen test substances for binding to a BRCA2 specific binding partner, to find mimetics of the BRCA2 polypeptide, e.g. for testing as cancer therapeutics.

Following identification of a substance which modulates or affects polypeptide activity, the substance may be investigated further. Furthermore, it may be manufactured and/or used in preparation, i.e. manufacture or formulation, of a composition such as a medicament, pharmaceutical composition or drug. These may be administered to individuals.

Thus, the present invention extends in various aspects not only to a substance identified using a nucleic acid molecule as a modulator of polypeptide activity, in accordance with what is disclosed herein, but also a pharmaceutical composition, medicament, drug or other composition comprising such a substance, a method comprising administration of such a composition to a patient, e.g. for treatment (which may include preventative treatment) of cancer, use of such a substance in manufacture of a composition for administration, e.g. for treatment of cancer, and a method of making a pharmaceutical composition comprising admixing such a substance with a pharmaceutically acceptable excipient, vehicle or carrier, and optionally other ingredients.

A substance identified using as a modulator of polypeptide function may be peptide or non-peptide in nature. Non-peptide "small molecules" are often preferred for many in vivo pharmaceutical uses. Accordingly, a mimetic or mimic of the substance (particularly if a peptide) may be designed for pharmaceutical use.

The designing of mimetics to a known pharmaceutically active compound

is a known approach to the development of pharmaceuticals based on a "lead" compound. This might be desirable where the active compound is difficult or expensive to synthesise or where it is unsuitable for a particular method of administration, e.g. peptides are unsuitable active agents for oral compositions as they tend to be quickly degraded by proteases in the alimentary canal. Mimetic design, synthesis and testing is generally used to avoid randomly screening large number of molecules for a target property.

There are several steps commonly taken in the design of a mimetic from a compound having a given target property. Firstly, the particular parts of the compound that are critical and/or important in determining the target property are determined. In the case of a peptide, this can be done by systematically varying the amino acid residues in the peptide, eg by substituting each residue in turn. Alanine scans of peptide are commonly used to refine such peptide motifs. These parts or residues constituting the active region of the compound are known as its "pharmacophore".

Once the pharmacophore has been found, its structure is modelled to according its physical properties, eg stereochemistry, bonding, size and/or charge, using data from a range of sources, e.g. spectroscopic techniques, X-ray diffraction data and NMR. Computational analysis, similarity mapping (which models the charge and/or volume of a pharmacophore, rather than the bonding between atoms) and other techniques can be used in this modelling process.

In a variant of this approach, the three-dimensional structure of the ligand and its binding partner are modelled. This can be especially useful where the ligand and/or binding partner change conformation on binding, allowing the model to take account of this in the design of the mimetic.

A template molecule is then selected onto which chemical groups which mimic the pharmacophore can be grafted. The template molecule and the chemical groups grafted on to it can conveniently be selected so that the mimetic is easy to synthesise, is likely to be pharmacologically acceptable, and does not degrade in vivo, while retaining the biological activity of the lead compound. Alternatively, where the mimetic is peptide based, further stability can be achieved by cyclising the peptide, increasing its rigidity. The mimetic or mimetics found by this approach can then be screened to see whether they have the target property, or to what extent they exhibit it. Further optimisation or modification can then be carried out to arrive at one or more final mimetics for in vivo or clinical testing.

Screening for Substances Affecting BRCA2 Expression

The present invention also provides the use of all or part of the nucleic acid sequence of the BRCA2 promoter region shown in figure 6 in methods of screening for substances which modulate the activity of the promoter and increase or decrease the level of BRCA2 expression.

"Promoter activity" is used to refer to ability to initiate transcription. The level of promoter activity is quantifiable for instance by assessment of the amount of mRNA produced by transcription from the promoter or by assessment of the amount of protein product produced by translation of mRNA produced by transcription from the promoter. The amount of a specific mRNA present in an expression system may be determined for example using specific oligonucleotides which are able to hybridise with the mRNA and which are labelled or may be used in a specific amplification reaction such as the polymerase chain reaction. Use of a reporter gene facilitates determination of promoter activity by reference to protein production.

Further provided by the present invention is a nucleic acid construct comprising a BRCA2 promoter region set out in figure 6 or a fragment, mutant, allele, derivative or variant thereof able to promoter transcription, operably linked to a heterologous gene, e.g. a coding sequence. A "heterologous" or "exogenous" gene is generally not a modified form of BRCA2 Generally, the gene may be transcribed into mRNA which may be translated into a peptide or polypeptide product which may be detected and preferably quantitated following expression. A gene whose encoded product may be assayed following expression is termed a "reporter gene", i.e. a gene which "reports" on promoter activity.

The reporter gene preferably encodes an enzyme which catalyses a reaction which produces a detectable signal, preferably a visually detectable signal, such as a coloured product. Many examples are known, including β-galactosidase and luciferase. β-galactosidase activity may be assayed by production of blue colour on substrate, the assay being by eye or by use of a spectrophotometer to measure absorbance. Fluorescence, for example that produced as a result of luciferase activity, may be quantitated using a spectrophotometer. Radioactive assays may be used, for instance using chloramphenicol acetyltransferase, which may also be used in non-radioactive assays. The presence and/or amount of gene product resulting from expression from the reporter gene may be determined using a molecule able to bind the product, such as an antibody or fragment thereof. The binding molecule may be labelled directly or indirectly using any standard

technique.

Those skilled in the art are well aware of a multitude of possible reporter genes and assay techniques which may be used to determine gene activity. Any suitable reporter/assay may be used and it should be appreciated that no particular choice is essential to or a limitation of the present invention.

Nucleic acid constructs comprising a promoter (as disclosed herein) and a heterologous gene (reporter) may be employed in screening for a substance able to modulate activity of the promoter. For therapeutic purposes, e.g. for treatment of cancer a substance able to up-regulate expression of the promoter directing the expression of normal may be sought. Alternatively, substances to down-regulate the promoter may help to prevent or inhibit the production of mutated BRCA2 polypeptide, if this is an agent implicated in the development of cancer. A method of screening for ability of a substance to modulate activity of a promoter may comprise contacting an expression system, such as a host cell, containing a nucleic acid construct as herein disclosed with a test or candidate substance and determining expression of the heterologous gene.

The level of expression in the presence of the test substance may be compared with the level of expression in the absence of the test substance. A difference in expression in the presence of the test substance indicates ability of the substance to modulate gene expression. An increase in expression of the heterologous gene compared with expression of another gene not linked to a promoter as disclosed herein indicates specificity of the substance for modulation of the promoter.

A promoter construct may be introduced into a cell line using any technique previously described to produce a stable cell line containing the reporter construct integrated into the genome. The cells may be grown and incubated with test compounds for varying times. The cells may be grown in 96 well plates to facilitate the analysis of large numbers of compounds. The cells may then be washed and the reporter gene expression analysed. For some reporters, such as luciferase the cells will be lysed then analysed.

Following identification of a substance which modulates or affects promoter activity, the substance may be investigated further. Furthermore, it may be manufactured and/or used in preparation, i.e. manufacture or formulation, of a composition such as a medicament, pharmaceutical composition or drug. These may be administered to

individuals.

Construction of Animal Models for BRCA2 Deficiency

The construction of animal models for BRCA2 deficiency can be carried out using standard techniques for introducing mutations into, for example, a mouse germ-line. In one example of this approach, a vector carrying an insertional mutation within exon 11 of the mouse BRCA2 gene is transfected into embryonic stem cells. Clones in which the mutant version of the gene has replaced with wild type are identified by Southern blot hybridisation. The clones are then amplified and cells are injected into mouse blastocyst stage embryos. Mice in which the injected cells have contributed to the development of the mouse are identified by Southern blotting. These chimeric mice are then bred to produce mice which carry one copy of the mutation in the germ line. These heterozygous mutant animals are then bred to produce mice carrying mutations in the BRCA2 gene homozygously. The mice having a heterozygous mutation in BRCA2 may be a suitable model for human individuals having one copy of the gene mutated in the germ line who are at high risk of developing breast cancer.

Example 1 Identification of the BRCA2 Gene

Following the definition of the interval D13S289-D13S267 in which the BRCA2 gene is believed to be located (Wooster et al (1994)), the following set of procedures were used identify the gene itself. A yeast artificial chromosome (YAC) contig was constructed of YACs believed to be in the region from the CEPH database. Chimerism of YACs was examined using fluorescent in situ hybridisation. Overlaps between YACs were established by amplification of sequence tagged sites (STS), hybridisation of STS PCR products and alu-PCR fingerprinting.

Using probes localised to and derived from the YAC contig, a P1 artificial chromosome (PAC) library was screened and positive PACs isolated. PACs were rehybridised to the PAC library to fill in gaps between clones. A PAC contig over a 1 megabase region was then assembled.

Overlaps between PACs were identified by:

(a) amplification of STS;
(b) hybridisation of end probes produced by linear PCR;
(c) HindIII/Sau3A fingerprinting.

Additional polymorphic markers from the region were identified by screening M13 libraries constructed from PACs, with oligonucleotides

containing repetitive sequences that are commonly polymorphic (GTn, GAn, CAGn, GATAn, GAATn). Using new markers, the region in which BRCA2 is likely to be located was narrowed further.

Transcripts from the region were identified by use of PAC DNAs in:

(a) exon trapping experiments using a lambda GET vector;

(b) hybrid selection between PAC genomic DNA and cDNA libraries.

Exons and cDNA fragments identified by these methods were sequenced.

Using primers synthesised from these sequences as probes, 15kb genomic subclones of the PACs that carry all or part of the cDNAs and exons were identified. These were sequenced from the cDNA sequence out into adjacent flanking genomic sequences.

Using oligonucleotide primers synthesised on the basis of these flanking genomic sequences, genomic DNA fragments containing the potential transcript were amplified and screened for mutations through the following DNA samples; DNAs from 46 families that show evidence of linkage to BRCA2 and/or show evidence against linkage to BRCA1 and/or do not have BRCA1 mutations and/or do have evidence of male breast cancer. Also screened were cancer cell lines that are homozygous for all polymorphic markers through the BRCA2 region; DNAs from 12 primary human tumour DNAs that show evidence of loss of heterozygosity in the BRCA2 region.

Sequence variants were detected by running ^{32}P labelled amplification products through:

(a) non denaturing polyacrylamide gels, at room temperature and at 4°C;

(b) denaturing polyacrylamide gels.

Fragments that showed altered mobility were reamplified using the PCR and directly sequenced. Sequences were run on an ABI 377 DNA sequencer.

In the course of this work, several sequence variants were detected, most of which are believed to be non disease associated polymorphisms. However, an abnormality detected in one fragment was predicted to destroy a splice site and create a termination codon in a family that shows strong linkage to BRCA2. The variant segregates with the disease chromosome. This is precisely the type of abnormality that would be expected within a cancer susceptibility gene and made this fragment a strong candidate for part of the BRCA2 gene.

This fragment was then used (a) as a probe in screening of cDNA libraries, (b) starting sequence for PCR amplification experiments from cDNAs and cDNA libraries. Positive clones and amplification fragments were sequenced. These results are shown in the figures 1 and 2.

From the initial isolation of this part of exon 16 of the BRCA2 gene, a variety of conventional techniques were then used to isolate the remaining portions of the BRCA2 sequence. The 900,000bp sequence released on the Internet (the Sanger sequence) assisted in this procedure. Figure 4 shows around 75% of the cDNA sequence isolated by the inventors, while the complete BRCA2 gene sequence, including the exon/intron structure, is shown in figure 7. The promoter region of the gene is set out in figure 6. These sequences could be downloaded from the ftp web sites to a local computer. The sequences can then be analysed using programs such as BLAST, FASTA, GRAIL or GeneFinder to identify putative coding regions. Oligos can then be designed to these predicted coding regions and these would be used in RT-PCR to confirm or refute the coding regions. The coding regions would also be compared with the experimental results obtained from the procedures set out herein.

Thus, given knowledge of a part of the BRCA2 sequence and its orientation within the Sanger sequence, the skilled person could readily isolate the rest of the sequence using a conventional exon prediction program to localise potential exons/open reading frames in the same orientation as the known portion of the BRCA2 sequence, and then do exon connection in which potential exons are used to design primers that sit in putative coding sequence. The primers could be designed in such a way as to have a primer based on known BRCA2 sequence and an unknown primer, such that when these are used to prime from cDNA, a product contiguous with known sequence would be produced if the unknown primer is a part of the BRCA2 gene. This process could be continued in an iterative manner, readily allowing the skilled person to walk through the Sanger sequence to obtain the full BRCA2 sequence, obtaining the intron/exon boundaries as part of the procedure. In the case of BRCA2, a large portion of the coding sequence (about 5kb) is in exon 11, part of which is included in the 10% of the sequence disclosed in figures 1 and 2.

By way of reference, the techniques described above and others useful in the isolation of genes by positional cloning are reviewed in Monaco, Curr. Opinion Gen. Devel., 4:360-365, 1994 and summarised in EP-A-0705902.

By way of further assistance, a protocol for the isolation of the full

length sequence and the polypeptide encoded by the BRCA2 gene from the sequence shown in figures 1, 2 and 4 is set out below, and could be used iteratively to isolate successive portions of the BRCA2 gene.

Screening cDNA libraries by hybridisation.

Presently identified cDNA fragments can be ^{32}P labelled and hybridised to various widely available plated or gridded cDNA libraries. Positive clones can then be isolated, and subject to replating and rehybridisation if necessary until a pure clone has been isolated. DNA can then be made from pure clones and will be sequenced by conventional Sanger dideoxy sequencing on a ABI 377 DNA sequencer.

Screening cDNA libraries by PCR amplification.

Oligonucleotides based on sequences within the BRCA2 sequences disclosed herein can be used in conjunction with oligonucleotides designed to prime from the cloning vector in PCR amplifications of aliquots of widely available cDNA libraries. This will allow amplification of fragments of the BRCA2 cDNA positioned between the currently known fragment and the cloning insertion site. Products of the PCR amplification can then be sequenced using Sanger dideoxy sequencing on an ABI 377 sequencer.

Rapid amplification of cDNA ends (RACE).

Primary cDNAs synthesised from a number of different tissue RNAs can be ligated to an oligonucleotide linker. After purification, PCR amplifications can be performed using an oligonucleotide that primes from the current BRCA2 cDNA sequence and a second oligonucleotide that primes from the linker. Amplification products will be directly sequenced using Sanger dideoxy sequencing. RACE is described in (6).

The new sequences can then be integrated into the full sequence of the gene by detection of overlaps with previously known components of the sequence.

The screening of cDNA or genomic libraries with selected probes can be conducted using standard procedures, for instance as described in Ausubel et al and Sambrook et al (supra).

These techniques allow the full coding sequence to be isolated based on the information disclosed in figures 1, 2 and 4. The full length sequence is defined as the sequence between a translation initiation codon (ATG) and a translation termination codon (TAA, TAG, TGA) between which there is an open reading frame. This in turn can be used to define the intron-exon structure of the gene. Primers can then be designed to flank each exon so that the whole coding sequence of the

gene can be amplified from genomic DNA, see for example the primers disclosed in figure 8 or in Nature Genetics, 12:333-337, 1996.

Further fragments of coding sequence were then amplified from genomic DNA in the previously described mutation testing set in order to detect additional disease associated mutations in the BRCA2 gene.

Example 2 Mutations in the BRCA2 Gene
Identification of Six Mutations in the BRCA2 Gene

In a first study, the inventors found a series of mutations in the BRCA2 sequence by comparing the native sequence with sequences obtained from families with a history of multiple cases of early onset breast cancer. The locations of these mutations in the amino acid sequence are shown by boxing the residues in the native sequence which are affected (see figure 5). The mutations to the BRCA2 gene are summarised in table 3, which shows the families in the left hand column against mutations in the BRCA2 gene in the right hand column. The remaining columns in table 3 from left to right are as follows:

Number of FBCs	=	Number of female breast cancers in the family.
Number of FBCs<50	=	Number of female breast cancers occurring in individuals under 50 years of age.
Number of OvCs	=	Number of ovarian cancers in the family.
Number of MBCs	=	Number of male breast cancers.

The columns referring to LOD scores at BRCA1 and BRCA2 show the chance that the incidence of the cancers in a family is linked to the BRCA1 or BRCA2 genes, with a more positive value indicating a greater chance of linkage. Thus, for example, a LOD score of +3 indicates a strong linkage between the incidence of cancer and the given gene in that family.

To identify BRCA2, genomic DNA fragments of less than 300bp containing putative coding sequences were screened for mutations. At least one affected member of 46 breast cancer families was examined. Each family included in this set either shows evidence of linkage to BRCA2, and/or shows evidence of breast cancer. The majority, but probably not all, of these families would be expected to be due to BRCA2 mutations.

Disease associated mutations in most known cancer susceptibility genes usually result in truncation of the encoded protein and inactivation of critical functions. In the course of the mutational screen of candidate coding sequences from the BRCA2 region, the first detected sequence variant that was predicted to disrupt translation of an encoded protein was observed in IARC 2932. This family is clearly linked to BRCA2 with a multipoint LOD score of 3.01 using D13S260 and D13S267. A six base pair deletion removes the last five bases of the exon examined (exon S66), deletes the conserved G of the 5' splice site of the intron, and directly converts the codon TTT for phenylalanine to the termination codon TAA. By sequencing this mutation has been detected in lymphocyte DNA from two other early onset breast cancer cases in this family. The individuals examined share only the disease-associated haplotype. The mutation is absent in more than 500 chromosomes from normal individuals or in the remaining families and cancers. This finding therefore identified the candidate gene which was proved by the work described herein to be the BRCA2 gene.

To characterise the BRCA2 gene further, exon S66 was used to isolate a series of cDNA clones which represented segments of the BRCA2 candidate. From alignment of the cDNA and genomic sequence data, the candidate BRCA2 gene was found to lie in three sequence contigs which also contained other previously isolated transcribed sequences. The exon and open reading frame prediction program Genemark was used to define putative additional 5' exons of the gene. Contiguity of the transcription unit was confirmed by RT-PCR on cDNA and sequence analysis. The availability of extensive sequence information at the cDNA and genomic level allowed mutational analysis of further coding regions of the putative BRCA2 gene in samples from breast cancer families.

A TG deletion (6819delTG) and a TT (6503delTT) deletion were detected in families CRC B196 and CRC B211 respectively (tables 1 and 2). In both families the mutation has been detected by sequencing other individuals with early onset breast cancer who share only the haplotype of 13q microsatellite markers that segregates with the disease.

Therefore, the mutations are on the disease associated chromosomes. A CT deletion was detected in family IARC 3594. This mutation has arisen within a short repetitive sequence (CTCTCT), a feature that is characteristic of deletion/insertion mutations in many genes and which is presumed to be due to slippage during DNA synthesis. Finally, a T deletion (6174delT) and an AAAC deletion have been found in Montreal 681 and 440 respectively. Both these families include a male breast cancer case and previous analyses have indicated that the large

majority of such families will have BRCA2 mutations. All these mutations are predicted to generate frame shifts leading to premature termination codons. None of the mutations have been found in chromosomes from over 500 healthy women and are therefore unlikely to be polymorphisms. The identification of several different germline mutations that truncate the encoded protein in breast cancer families that are highly likely to be due to BRCA2 strongly suggests that we have identified the BRCA2 gene. In particular, the 6174delT mutation from Ashkenazi Jewish family data is reproduced in other families examined and so may be useful in screening individuals for a susceptibility to cancer, especially male or female breast cancer or ovarian cancer.

Northern analysis has demonstrated that BRCA2 is encoded by a transcript of 10-12kb which is present in normal breast epithelial cells, placenta, and the breast cancer cell line MCF7. This suggests that our present contig of cDNAs covering approximately 9kb (including 1.6kb of 3' untranslated sequence) may not include the whole BRCA2 coding sequence. The known sequence of 2329 amino acids encoded by the BRCA2 gene does not show strong homology to sequences in the publicly available DNA or protein databases. The homology and motifs of BRCA2 found in an analysis of the corresponding genes in other species is described in example 6. However, some weak matches were detected including, intriguingly, a very weak similarity to the BRCA1 protein over a restricted region (amino acids 1394-1474 in BRCA1 and 1783-1863 in the portion of BRCA2 shown in figure 5).

Identification of Further Mutations in the BRCA2 Gene

Patient material

Families were identified which contained a minimum of either three cases of breast cancer and for which DNA samples were available from a least one affected individual. Complete or partial analysis of the BRCA1 gene identified those families with disease associated germline mutations. The remaining families were analysed throughout the complete BRCA2 coding sequence for germline mutations, which are shown in table 1 against the incidences of breast cancer (including early onset breast cancer) and ovarian cancer. Borderline ovarian cancer was not included. Confirmation of diagnosis was available from pathology reports or death certificates.

Mutation analysis of BRCA2 in ovarian cancer families

Seventy seven families with two or more first degree relatives with epithelial ovarian cancer were analysed for BRCA2 mutations. Mutation screening was performed using a combination of the protein truncation

test (PTT) and non-radioactive single-strand conformation analysis/heteroduplex analysis (SSCA/HA). PTT was performed from genomic DNA for exon 11 in all cases and for the entire coding region in those individuals in which RNA derived from lymphoblastoid cell lines was available to perform reverse-transcriptase PCR. Primers were designed to PCR amplify exon 11 or the complete coding sequence in overlapping fragments ranging in size from 10 to 1.3kb.

PTT was performed using the TNT rabbit reticulocyte lysate system (Promega) incorporating ^{35}S methionine (Amersham) for protein detection. Protein products were electrophoresed on 12-15% SDS polyacrylamide gels at 30-60mA for 12-16 hours. SDS-PAGE gels were fixed in 30% methanol/10% glacial acetic acid, dried and exposed to Kodak X-Omat film for 16-72 hours. The approximate location of the sequence alteration resulting in truncated protein variants was localised and the region sequenced to confirm the precise nucleotide alteration. SSCA/HA was performed on genomic DNA for coding exons 2-10 and 12-27. The 5' and 3' splice boundaries for exon 11 were also analysed. SSCA and HA variant conformers were sequenced as previously described (3) in order to characterise the precise nucleotide change. The ability to detect mutations identified by PTT using SSCA/HA was detectable using SSCA/HA on fragments ranging in size from 300-600bp. All primer sequences designed for PTT and SSCA/HA are available on request (e-mail sg200@cam.ac.uk).

Mutation analysis of BRCA2 in breast cancer families

Statistical methods

We tested for an association between mutation location and disease phenotype using a permutation argument similar to that used by Gayther et al (12). Test statistics were based on standard chi-squared statistics for the difference in the rate of ovarian cancers, as a proportion of all ovarian and breast cancers, confirmed by position. In the main analysis, the statistic X was calculated assuming an alternative hypothesis in which different rates applied for mutations in three different regions, i.e. before codon n1-n2 and after n2 (i.e. a 2 degree of freedom chi-squared). The three putative missense mutations were ignored in this analysis as their significance is unclear. The statistic X was maximised over all possible values of the cutpoints n1 and n2 giving the statistic Xm. The significance of this maximised chi-squared was then computed empirically by calculating the value of Xm in 50,000 datasets in which the 26 mutations were randomly permuted among the families. We also computed a statistic to test for a trend in the ratio of ovarian:breast cancer risk along the gene, based on a chi-squared test for trend, as in (12) which was not

significant.

In an attempt to confirm the association using other data from previously published reports, we computed the chi-squared statistic (two degrees of freedom) using data from five studies (9,10,15-17), with the cutpoints n1 and n2 fixed at the best values for the UK dataset. The significance of this result was then based on permuting the mutations among families as before.

Discussion

BRCA2 probably accounts for the majority of high risk families not accounted for by BRCA1. The identification of BRCA2 should, therefore allow more comprehensive evaluation of families at high risk of developing breast cancer. However, the role of environmental, lifestyle or genetic factors in modifying the risks of cancer in gene carriers is unknown and further studies will be required before routine diagnosis of carrier status can be considered.

Although, many of the mutations described in these examples are not is common to the BRCA2 genes of different families, these mutations may nevertheless prove to be a useful in methods of diagnosing or detecting a predisposition to cancer, especially breast cancer. In addition, on further routine analysis of a larger sample of families, it may emerge that a given mutation is common to a significant portion of cancer cases and so might form the basis of a simple diagnostic test. Even if this is not the case, knowing the location of the mutations may help to reduce the amount of sequencing work that would be needed in carrying out a test on a patient, e.g. by sequencing all or part of their BRCA2 gene, thereby to determine a predisposition or susceptibility to cancer. Further applications and uses of the mutations are set out in the general section above.

Germline mutations of BRCA2 are predicted to cause approximately 35% of families with multiple case, early onset female breast cancer, and they are also associated with an increased risk of male breast cancer, ovarian cancer, prostrate cancer and pancreatic cancer (5,9,10). Germline mutations of a second cancer susceptibility gene BRCA1 (3), are associated with a strong predisposition to ovarian cancer as well as female breast cancer (11). Recent studies have suggested that the phenotype in BRCA1 families with respect to the ratio of breast to ovarian cancer varies with the location of the BRCA1 mutation (12,13). To determine whether germline mutations in BRCA2 are associated with a similar variation in phenotypic risk, we have analysed the distribution of mutations of families with multiple cases of breast and/or ovarian

cancer ascertained in the United Kingdom and Eire. The majority of these mutations lead to premature truncation of BRCA2 as a result of frameshift deletions, insertions, nonsense mutations or splice site alterations. Analysis of the mutation distribution along the length of the gene indicates a significant genotype-phenotype correlation. Truncating mutations in families with the highest risk of ovarian cancer relative to breast cancer are clustered in a region of approximately 3.3kb in exon 11 (p=0.0005). Published data on mutations in 45 other BRCA2 linked families provide support for this correlation.

Families were ascertained on the basis of either three or more cases of breast cancer or two or more first degree relatives with epithelial ovarian cancer diagnosed at any age. Mutations of the BRCA1 gene were then identified and genomic DNA of a single affected individual from each of the remaining families was used to screen for mutations throughout the coding sequence of BRCA2 using a combination of single-strand conformation polymorphism (SSCP) and protein truncation (PTT) assays. The BRCA 2 coding sequence consists of 10,248 nucleotides encoded by 26 exons (15). The majority of exons are relatively small although exons 10 and 11 represent approximately 60% of the entire coding region. The mutations found are set out in tables 1 and 2.. The mutation spectrum consisted of 22 frameshift deletions or insertions, 3 nonsense mutations and one splice site alteration. These mutations are all predicted to result in premature truncation of the predicted BRCA2 peptide. Three novel putative missense mutations were also detected. Mutations were distributed throughout the gene and only one additional mutation, 6503delTT, was found to be recurrent.

BRCA2 mutation data from families ascertained outside the UK derived from previous reports and unpublished data from our own laboratories provide support for this clustering. BRCA2 mutations have been identified in 45 such families (9,10,15-17). The 17 families with mutations in the OCCR are reported to contain 11 ovarian and 45 breast cancer cases compared to 22 ovarian and 282 breast cancer cases in the remaining families (odds ratio 22.9; permutation test for differences p=0.007). Consistent with this, there are three large BRCA2 families where reasonably systematic evidence on cancer risks are available, namely UTAH 107 (15), CRC 186 and the Icelandic family (10). Mutations in all three families are located near the 5' or 3' end of the gene and appear to be associated with a high lifetime risk of breast cancer, with mutations in the OCCR being associated with a higher ovarian cancer risk than average, or a lower breast cancer risk, or both. In this regard, future data on the 6174delT mutation will be particularly important. This mutation, which lies in the OCCR, is common in Ashkenazi Jews and direct population based estimates of its prevalence

in breast and ovarian cancer patients should be possible.

The observed genotype-phenotype correlation is somewhat surprising given that, unlike in BRCA1, a region in the centre of the gene appears to be associated with a distinct risk. There are precedents in cancer susceptibility genes, for example the adenomatous polyposis coli gene, where regions of mutation clustering within the gene are associated with predisposition to a specific phenotype (18,19). We do not yet know why truncating mutations in a central portion of the BRCA2 gene should result in increased risk to ovarian cancer. No regions of functional homology between BRCA2 and other genes have been identified. A series of eight internal amino acid repeats have recently been observed in exon 11 but these have little homology to other known genes (20). It is perhaps interesting that all eight repeats are contained within the OCCR. That the mutations cluster in a single exon suggests an explanation based on alternative splicing. Complete or partial splicing of exon 11 may produce alternatively spliced forms with the ability to "rescue" mutations in breast but not ovarian epithelium. Our results suggest that they may be interesting differences in the structure or function of BRCA2 between breast or ovarian epithelium.

Example 3 Expression of BRCA2 Polypeptide

Construction of Full-Length cDNA Clones for Human BRCA2

The approximately 10.5kb cDNA coding for human BRCA2 was assembled from fragments of cDNA amplified by PCR and from cloned genomic DNA. All residue numbers are from Genbank accession number U43746.

The 5' end of the cDNA (residues 204-2357) was amplified by PCR from cDNA made from RNA from MCF7 cells using the primers:

CATTGGAGGAATATCGTAGG (bases 204-223) and
GACAGAGAATCAGCTTCTGG (bases 2338-2357).

This 2.1kb fragment was cloned into pBluescript and the sequence determined.

This plasmid (p12302) was then digested with Asp718 which cuts in the plasmid polylinker and partially with BamHI which cuts as bases 238 and 792. The double-stranded oligonucleotide:

GTACCGCCGCCATGGAACAGAAGATTTCCGAAGAAGATCTGCCTATTG
GCGGCGGTACCTTGTCTTCTAAAGGCTTCTTCTAGACGGATAACCTAG

was then ligated into the plasmid and recombinants carrying the oligo

ligated to the plasmid digested at base 238 were selected.

This plasmid was digested with NdeI which cuts at bp 1795 and SmaI which cuts in the polylinker of the plasmid. This was ligated in a three-way ligation to a PCR product (1795-2280) which had been digested with NdeI and BstY1 and residues 2280-6571 as a BstY1-BsiHKAI fragment. The latter fragment consists of exon 11 of the BRCA2 gene and was isolated from a plasmid carrying genomic DNA. This generated plasmid p12672.

The 3' end of the BRCA2 cDNA was generated by RT-PCR with primers:

AAAAGTAACGAACATTCAGACCA (bases 6277-6299) and
ATTGTCGCCTTTGCAAATGC (bases 10,486-10,505) on cDNA from MCF7 cells.

This fragment was subcloned into pBluescript to generate plasmid p12661. Subsequently this fragment was amplified with a primer(GTACTCCAGAACATTTAATATCC bases 6325-6344) and the T3 primer from bluescript.

This fragment digested with Bsp1201 was then ligated into SnaBI-NotI cut p12672 to generate the full-length clone p12806.

This plasmid was subsequently modified to include a FLAG epitope as follows. p12806 was digested with Asp718 and SalI and then ligated to pBlueBac2B digested with the same enzymes. This generated plasmid p13013. This was digested with SacI and BgIII and ligated to a linker coding for a consensus Kozak sequence and a Flag epitope as detailed below:

CGGGTACCAGATCTGCCGCCACCATGGATTACAAGGACGACGATGACAAG
TCGAGCCCATGGTCTAGACGGCGGTGGTACCTAATGTTCCTGCTGCTACTGTTCCTAG

This full-length BRCA2 cDNA was ligated into expression vectors. In particular a modified version of the vector pMTSM was used this vector carries an adenovirus major late promoter and SV40 origin of replication.

Expression of BRCA2 in Mammalian Cells

The full-length BRCA2 expression vector was transfected into COS cells using standard techniques. The expression of the protein was monitored by Western blotting immunoprecipitation and immunofluorescence. This demonstrated that the BRCA2 protein could be detected as an approximately 400kD protein similar in size to that predicted by the sequence of the cDNA. Under these conditions, in this cell-type, the

protein had a complex subcellular localisation being located in either a membrane-like compartment (most probably the endoplasmic reticulum) or in the nucleus or in both (see figures 11 to 13).

Interaction of BRCA2 With Other Proteins

This can be assessed using a yeast two-hybrid system to clone proteins which interact with BRCA2. As this procedure is potentially complicated by the size of the BRCA2 ORF, the ORF may be divided up into 5-10 fragments and screened separately. These fragments will be used as 'bait' both episomally and as integrants into the yeast genome and used to screen peptide libraries such as those derived from HeLa cells, human fetal brain and a new library derived from normal human breast. Clones confirmed as interacting in the two hybrid system will assayed for interaction both in vivo and in vitro. Bacterially expressed GEX-fusion proteins will be tested for direct interaction in vitro. Epitope-tagged versions of BRCA2 and the potential interacting proteins will be co-microinjected into cell lines and their sub-cellular location determined by immunofluorescence to establish co-localization. Co-transfection and cross-immunoprecipitation could be used to establish that the two proteins interact in vivo. In addition to identifying novel BRCA2 interacting proteins, the above approaches may be used to ascertain whether BRCA2 can dimerise or interact directly with BRCA1.

The nature of BRCA2 interacting proteins can also be directly determined by biochemical fractionation followed by mass spectroscopy. Immunoprecipitation of ^{35}S-labelled extracts followed by SDS-gel electrophoresis can be used produce molecular weight estimates of these proteins. These would be further analysed by analytical 2D gel electrophoresis followed by MALDI-TOF mass-spectroscopy and peptide mass-mapping (23). This technique allows the certain identification of proteins whose sequences are present in the databases and assignment of likely family members (>80% identity).

Biological Function of BRCA2

An experiment can be performed to test whether BRCA2 expression can block the growth of breast and ovarian cancer cell lines specifically. Ideally such experiments make use of breast tumour cells that do not express BRCA2, which can readily be identified by screening existing breast cancer cell lines for absence of BRCA2 expression. Alternatively, cell lines could be established from patients shown to lack wild type BRCA2. It is also possible that over-expression of BRCA2 in cancer cells that still express will also suppress growth.

BRCA2 cDNAs can suitably be expressed under the control retroviral LTR

(pBABE) or elongation factor 1α promoters (the pEFBos series - (22)). Plasmids can be co-transfected with drug resistance markers and the number of colonies that grow out compared to vector controls. surviving colonies can be expanded and tested for tumourgenecity by injection into nude mice. In instances where the inhibitory action of the proteins cannot be detected in the relatively long term colony growth assay as the transfected plasmids do not stably express, microinjection of BRCA2 expression constructs can be used whereby the injected cells are detected by immunocytochemistry of the exogenous protein or by co-injection of a marker plasmid (21). In addition, time-lapse video recording, as described above, could be used to determine whether any growth inhibition effect is cell autonomous, i.e. whether the effect is paracrine or autocrine. These systems will also be useful for detecting any apoptotic effect.

Example 5 **Methods for Detecting BRCA2 Mutations**

Migration shift assays to detect BRCA2 mutations.

DNA amplification in the PCR.

25ng of genomic DNA from each individual to be screened for mutations is amplified in 35 cycles of the PCR using the pertinent oligonucleotide primers (see primer list). Prior to incorporation into the PCR, both oligonucleotide primers are end radiolabelled with gamma ^{32}P using T4 polynucleotide kinase. Following amplification in the PCR, formamide loading dye is added to each sample and the sample denatured at 94°C for 3 minutes. Following denaturation the sample is placed immediately on ice.

DNA fragment sizing.

2μl of each sample is loaded immediately onto a well formed by a 40 slot sharks' tooth comb in conventional 0.4mm thick denaturing 6% polyacrylamide gel. The sample is electrophoresed through the gel for 2-5 hours at 90 Watts at room temperature.

SSCP heteroduplex analysis

SSCP is a PCR based assay for screening DNA fragments for sequence variants/mutations. It involves amplifying radiolabelled 100-300 bp fragments of the BRCA2 gene, diluting these products and denaturing at 95°C. The fragments are quick-cooled on ice so that the DNA remains in single stranded form. These single stranded fragments of BRCA2 are run through acrylamide based gels. Differences in the sequence composition will cause the single stranded molecules to adopt difference conformations in this gel matrix making their mobility different from wild type fragments, thus allowing detecting of mutations in the

fragments being analysed relative to a control fragment upon exposure of the gel to X-ray film.

These fragments with altered mobility/conformations are directly excised from the gel and directly sequenced for the mutation. Following denaturation the sample is cooled on ice for 10 minutes to allow the heteroduplex to form. Each sample is electrophoresed through two different types of gel.

A typical set of conditions for SSCP analysis are as follows. 3μl are electrophoresed overnight at 4 Watts at room temperature through a 6% non denaturing polyacrylamide gel containing 10% glycerol.

3μl are electrophoresed for four hours at 30 Watts in a 4°C cold room through a 4.5% non denaturing polyacrylamide acrylamide gel without glycerol.

Following electrophoresis, gels are dried onto Whatman 3MM paper, and placed in an autoradiography cassette at room temperature for a period ranging from two hours to several days.

Following development of the autoradiograph band shifts in individual samples are detected by eye.

Sequencing of PCR product.

Where a band shift is seen in SSCP heteroduplex or DNA fragment sizing gels, the fragment concerned is reamplified from the relevant stock genomic DNA and directly sequenced. To sequence PCR product, the product is precipitated with isopropanol, resuspended and sequenced using TaqFS+ Dye terminator sequencing kit. Extension products are electrophoresed on an ABI 377 DNA sequencer and data analysed using Sequence Navigator software.

BRCA2 PTT Assay

PTT is another PCR based screening assay. Fragments of BRCA2 are amplified with primers that contain the consensus Kozak initiation sequences and a T7 RNA polymerase promoter. These extra sequences are incorporated into the 5' primer such that they are in frame with the native coding sequence of the fragment being analysed. These PCR products are introduced into a coupled transcription/translation system. This reaction is allows the production of BRCA2 RNA from the fragment and translation of this RNA into a BRCA2 protein fragment. PCR products from controls make a protein product of a wild type size relative to the size of the fragment being analysed. If the PCR product analysed has a frame-shift or nonsense mutation, the assay will

yield a truncated protein product relative to controls. The size of the truncated product is related to the position of the mutation, and the relative region of the BRCA2 gene from this patient is sequenced to identify the truncating mutation.

The following protocol was adapted for a BRCA2 PTT assay. The PTT primers are shown in figure 10.

Each PTT primer is preceded by the T7/Kozak sequence
GGATCCTAATACGACTCACTATAGGGAGACCACCATG

1. Thirty five cycle primary PCR reaction in 20ul.
2. Product confirmation by electrophoresis on 2% agarose gel.
3. Three μl aliquot amplification with nested PTT prime for 15 cycles.
4. Product confirmation by electrophoresis on 2% agarose gel.
5. In vitro transcription/translation using Promega (CA) TNT kit, incorporating ^{35}S radiolabelled methionine.
6. Laemmli buffer reaction stop and denaturation.
7. Gel electrophoresis of product on 15% acrylamide gel at 16mA.
8. Fix, amplify and dry gel.
9. Autoradiographic exposure for 2 hours.

These approaches can be combined to provide an accurate and effective screen, in terms of results achieved, the economical cost and the time taken to provide the results. By way of example, a combined protocol used by the inventors involves the following:

1) DNA samples from familial reference cases (probands) are first screened via PTT. Exons 10, 11, and the terminal exon 27 are analysed using this technique. Exons 10 and 27 are done in 1 fragment, whilst the larger exon 11 is done in 2 fragments. If protein truncations are seen, the corresponding genomic region is sequenced from the patient to identify the exact mutation. This approach is able to screen approximately 60% of the coding region in rapid fashion.

2) Samples negative for the PTT screening are then analysed using SSCP. The entire coding region, including those regions already examined via PTT, is screened for mutations using SSCP. The coding sequence is amplified from genomic DNA using primer sets as described in this application. In addition, radiolabelled PCR products generated for SSCP analysis are also run on denaturing acrylamide sequencing gels which allows for detection of small size changes relative to control fragments, indicative of insertions or deletions. Any mutant fragments seen are excised directly from the gels and direct sequenced along with

the matching genomic DNA fragment to determine the exact mutation

Examples of screening carried out using these methods are set out in the table below.

BRCA2 screening in breast cancer family probands:

Family	Mutation	No. Female Br Ca	Exon	Method of Detection
M107	2810insA	8	11	SSCP/PTT, mutation detected on both assays
M742	259insT	6	2	SSCP
M688	594delCT	2 (2 Ov)	11	PTT
M347	1003delAG	3	9	SSCP
D-MBC2	1485delT	5 (1 Mbr)	10	PTT
M269	8764delAG	19	20	SSCP
M612	8764delAG	4	20	SSCP
M133	2034insA	11	10	SSCP, too close to end of exon to be reliably up by PTT

Example 6 Results of Sequence Alignment

Materials and Methods

Total genomic DNA was obtained for green monkey (*Ceropithecus aethiops*), hamster (*Critetulus griseus*), pig (*Sus scrofa*), dog (*Canis familiaris*), cow, chimpanzee, chicken, snake, zebra fish, *X.laevis*, *S.pombe* and *S.cerevisiae*. PCR primers were designed from a consensus sequence for the human BRC motif. These included primers that were an exact match to the human DNA sequence and degenerate primers based on the BRCA2 protein sequence. The sequences were as follows.

Motif forward set 1: pure primer -AAAGCTGTGAAACTGTT and a degenerate set which was an equal mixture of -AA(AG)GCIIIIAA(AG)CTITT and -AA(AG)GCIIIIAA(AG)TT(AG)TT where I is inosine.

Motif forward set 2: pure primer as for set 1 and a degenerate set made with -AA(AG)GCIGTIAA(AG)CTITT and -AA(AG)GCIGTIAA(AG)TT(AG)TT.

Motif reverse set: pure primer - TTCCCACTTGCAGTCTGAAA and a mixed set of -TICC(GA)CTIGCIGT(CT)TG(GA)AA and -TTICCIGAIGCIGT(CT)TG(GA)AA.

Further sets of primers were designed to a region of BRCA2 that is conserved between human and mouse. These were a pure primer of

AGCAAGCAATTTCAAGG and a degenerate set of - TCIAA(AG)CA(AG)TT(TC)GA(AG)GG and -AG(TC)AA(AG)CA(AG)TT(TC)GA(AG)GG. The PCR was performed using each forward primer in conjunction with a relevant reverse primer. The reaction conditions were as described in (4) except for the cycling conditions of the PCR. These were 94°C 1 mix X°C for 1 min. 72 C for 1 min, where X was °65 C for the first two cycles and then decreased by 2°C every two cycles until it reached either 55, 49, 45, 39 or 35°C, at which point X remained constant. A further 30 cycles were performed at the final annealing temperature. PCR products were resolved on agarose gels and discrete bands were excised and sequenced using a Taq dideoxy termination protocol (Perkin-Elmer). Sequence products were separated on an ABI377 automated DNA sequencer.

All of the sequencing reactions were performed at least twice and on both strands of the template. Species-specific PCR primers were designed to exon 11 of the BRCA2 gene using the above DNA sequence. These were used in conjunction with a set of 52 pairs of PCR primers designed to the human BRCA2 gene to amplify and sequence further segments of exon 11 from the monkey, dog, pig and hamster. A human BRCA2 probe (exon 11) was used to identify a λ clone containing a portion of the mouse BRCA2 gene. This was used to obtain a mouse-specific BRCA2 sequence that was in turn used to design mouse-specific PCR primers. These were used to identify positive clones in a mouse BAC library. Fragments of positive BACs were cloned, at random, and those shown to contain fragments of exon 11 of the BRCA2 gene were sequenced.

Sequencing the BRC motifs

Two approaches were used to obtain the sequence of BRCA2 exon 11 (including the BRC motifs) in five mammalian species, using the sequence of human BRCA2 was taken from earlier work, see also figure 7. The sequence of exon 11 of the mouse BRCA2 gene was obtained from a BAC clone isolated by low stringency hybridisation with fragments of the human BRCA2 gene. Fragments of BAC were cloned at random and those shown to contain fragments of the BRCA2 gene were sequenced. PCR primers were then designed to the BRC motifs and to regions that were conserved between human and mouse. These were used to amplify fragments of the BRCA2 gene in DNAs from 12 species (as described in Materials and Methods). A sequence showing similarity to human BRCA2 was obtained from green monkey, hamster, pig, dog, cow and chimpanzee. The sequences obtained from chicken, snake, zebra fish, *Xenopus laevis*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* showed no similarity to any BRCA2 sequence. Species-specific PCR primers were designed for green monkey, hamster, pig and dog. These were used in

combination with all of the PCR primer pairs used to amplify human BRCA2 exon 11 to amplify further segments of the exon. The process was repeated until all of the sequences between and including BRC motifs one and eight had been obtained for the green monkey, hamster and dog. Using this approach, it was only possible to obtain the sequence for repeats 3-8 for the pig. In total, 46 BRC repeats were sequenced in six different mammals (including human).

The BRC motif is conserved in mammalian species

The percentage identity of the translation of exon 11 of the BRCA2 gene (approximately half of the whole coding sequence) between the six mammals is shown in Table 1. Overall, the degree of conservation is low, with 58% identity between the 1602 residues of the human and mouse, 54% identity between mouse or hamster and dog and 49% between pig and mouse (over 928 amino acids). Even between closely related species such as human and monkey, and mouse and hamster, amino acid identity is only 93 and 72% respectively. However, an alignment of the sequences from the six mammals studied demonstrates that exon 11 translations can be aligned along their whole lengths and that the degree of conservation is variable (figure 9). There are a number of short regions of high identity. Some of these regions coincide with BRC motifs, for example BRC1, 2, 3, 4, 7 and 8 (figure 9). There are other highly conserved segments, for example amino acids 469-493 (using the numbering of the human sequence in figure 9) that show no similarity to the BRC motifs or anything else reported in the databases. None of the latter are repeated within BRCA2.

An alignment of all 46 motifs sequenced demonstrates the high degree of interspecies and intraspecies conservation between BRC1, 3, 4, 7 and 8. From this alignment, we have identified a region of 26 amino acids that is conserved in all of the BRC motifs (Fig. 2) which has allowed us to generate a BRC consensus sequence. It is possible to align some residues outside this common region; however, such alignments are not robust, being very sensitive to the parameters used. There are motifs that contain all of the consensus sequence, for example BRC4, while others, for example BRC6, show a considerable divergence from the consensus. In total, 30/46 of the BRC motifs (65%) have 11 or more of the 13 consensus residues while 87% have eight or more of the consensus sequence.

The following table shows the percentage identities between the translation of exon 11 of the BRCA2 gene from human, monkey, pig, dog, hamster and mouse.

	Human	Monkey	Pig	Dog	Hamster	Mouse
Human	100	93	62	70	57	58
Monkey		100	61	68	58	58
Pig			100	64	50	49
Dog				100	54	54
Hamster					100	72
Mouse						100

Discussion

The low sequence identities for exon 11 (shown in the table above) suggest that BRCA2 is evolving at a faster rate than most other cancer susceptibility/tumour suppressor genes. For example, there is 98% identity between human and mouse NF1 with 95 and 91% between human and mouse WT1 and Rb1 respectively. A notable exception, however, is BRCA1 which only shows 58% amino acid identity between mouse and human. Therefore, although BRCA1 and BRCA2 do not show substantial sequence similarity, they are similar by virtue of a high rate of evolution, unusual gene structure (both have a large exon 11, see (3) and (7) and lack of somatic mutations in sporadic cancers. Whether one or more of these features relates to a function congruence remains to be elucidated.

Although the identity for the translation of exon 11 of the BRCA2 gene is low, most of the BRC motifs are highly conserved between the species analysed. This suggests that there has been pressure to maintain the BRC repeats in BRCA2 and, therefore, that they are important in its function. However, from our alignment, BRC6 is much less conserved than BRC1, 2, 3, 4, 7 and 8 (31 and 85% identity between human and mouse BRC6 and BRC7 respectively). Moreover, it has been altered by insertions or deletions such that the length of the motif differs between species. BRC3 and BRC5 also shows less conservation than BRC1, 2, 3, 4, 7 and 8 (62 and 58% identity respectively between human and mouse BRC3 and BRC5). However, both exhibit a higher level of conservation than the exon 11 translation overall. The data suggests that multiplication of the motif took place prior to the mammalian radiation. For instance, several amino acid residues within motif units (especially in BRC2) are different from the equivalent residues in other units, but are highly conserved in mammalian species. Moreover, the sequences flanking the motifs are not conserved between repeats, but in some cases are conserved across species (for example BRC1 and 4, figure 9).

The PCR products that we derived from non-mammalian DNA did not exhibit any similarity to mammalian BRCA2. This suggests that either BRCA2 is restricted to mammals or that the non-mammalian orthologues of BRCA2 have diverged to such an extent that they cannot be identified by the techniques that we used. However, there is evidence of a weak similarity between the BRC sequence and a *Caenorhabditis elegans* gene (CET07E3_2), suggesting that the BRC motif is not restricted to mammals.

There are few clues to the function of the BRC sequences. Truncating germline mutations in the BRCA2 gene that predispose to the development of breast cancer are located throughout the coding region of the gene. In many cases, the mutations leave all the BRC repeats intact, giving little information on the relationship between the motifs and their role in the normal function of BRCA2. The spacing between individual motifs varies from -60 to 300 amino acids, but is reasonably well conserved between mammals. Furthermore, there are multiple elements of secondary structure within each motif which may indicate that they form globular domains. However, it is unclear how these may function. Therefore, direct investigation will be necessary to elucidate the biochemical functions of the BRC motifs.

Conserved regions such as the BRC motifs identified in exon 11 provide valuable indications of domains of the BRCA2 polypeptide that are likely to be important in determining its activity. Thus, these motifs are good candidates for screening studies described above, to find mimetics for BRCA2 polypeptide.

Motifs from exon 11 that are conserved between some or all of the species examined and depicted in figure 9 and are summarised in table 4:

Table 4

Start Amino Acid	End Amino Acid	Motif
379	404	SFRTASNKEIKLSEHNIKKSKMFFKD
594	619	GFYSALGTKLNVSSEALQKAVKLFSD
806	831	SFQTASGKNIRVSKESLNKAVNFFDQ
909	934	GFHTASGKKVKIAKESLDKVKNLFDE
1057	1082	AFYTGHSRKTSVSEASLLEAKKWLRE
1257	1271	AFSTASGKIVFVSHE
1392	1417	IFSTASGKSVQVSDASLQKARQVFSE
1474	1499	GFSTASGKQVSVSESALHKVKGMLEE

Table 1. Germline mutations of the BRCA2 gene in breast and ovarian cancer families

Family	Number of cases of cancer			Description of mutation				
	Breast	Breast ≤ 60 yrs	Ovarian	Exon	Codon	Nucleotide[1]	Alteration	Mutation Type#
B061	2	2	0	2	9	253	253delC	FS
B091	4	4	0	7	194	809 G>T	Trp194ter	N
B135	5	3	0	10	397	1418	1418ins4	FS
B105	2	2	0	10	434	1529	1529del4	FS
B022	4	3	0	10	598	2022	2022del5	FS
B169	3	3	0	11	643	2157	2157delG	FS
B124	3	3	0	11	936	3034	3034del4	FS
B211	4	4	4	11	2092	6503	6503delTT	FS
B150	6	5	0	11	2092	6503	6503delTT	FS
B160	4	2	0	11	2134	6630	6630del5	FS
B245	7	3	0	11	2154	6690	6690delTC	FS
B196	18	14	0	11	2197	6819	6819delTG	FS
B097	4	3	0	11	2277	7057	7057del5	FS
B007[d]	11	5	0	17	2637	8138	8138del5	FS
B028[d]	3	3	0	17	2654	8162	8162delG	FS
B186	18	18	1	22	2984	9179C>G	Ser2984ter	N
B247	3	3	0	23	2994	9208	9208ins4	FS
B001	4	3	0	25	3167	9729G>A	9729G>A	SS
M681[d]	3	3	0	11	1982	6174	6174delT	FS
M440[d]	2	2	0	11	936	3034	3034del4	FS

[d]These families also contain one or two male breast cancer cases. #Mutation types characterised as frameshift deletions or insertion (FS), nonsense mutations (N), missense mutations (M) or splice site alterations (SS).

Table 1 (cont). Germline mutations of the BRCA2 gene in breast and ovarian cancer families

Family	Number of cases of cancer			Description of mutation			
	Breast	Breast ≤ 50 yrs	Ovarian	Exon	Nucleotide	Alteration	Mutation Type#
M107	8	6	0	11	2810	2810insA	FS
M133	11	4	1	10	2034	2034insA	FS
M174	9	8	0	11	3034	3034del4	FS
M269[d]	21	11	0	20	8764	8764delAG	FS
M280	4	0	0	11	6085	G6085Tns	N
M284	3	2	1	11	4304	4304delC	FS
M347[d]	4	0	0	9	1003	1003delAG	FS
M372	6	0	0	25	9661	9661delGT	FS
M609	12[d]	0	0	10	2024	2024del5	FS
M612	4	3	0	20	8764	8764delAG	FS
M688	2	0	2	11	5946	5946delCT	FS
M733	2	0	0	11	6437	6437del4	FS
M742	6	4	0	2	259	259insT	FS
W9052	4	3	0	10	2034	2034delA	FS
W9116	4	3	0	10	1917	G1917Ans	N
DMBC2	6[d]	2	0	10	1485	1485delT	FS
W9012	1[d]	0	0	11	5839	5839del4	FS

[d]These families also contain one or two male breast cancer cases. #Mutation types characterised as frameshift deletions or insertion (FS), nonsense mutations (N), missense mutations (M) or splice site alterations (SS).

TABLE 2

Family	Mutation description	Mutation type	sub sequence	d/i sequence	Location exon	Location fragment	Location base	Location codon
CRC B001	Splice	s	G->A		25	B	9729	0
CRC B007	Frameshift	d5		CCTTT	17		8138	2637
CRC B022	Frameshift	d5		ATCTT	10	E	2022	598
CRC B028	Frameshift	d1		G	17		8162	2645
CRC B061	Frameshift	d1		C	2		253	9
CRC B097	Frameshift	d5		CTTAT	11	Z	7057	2277
CRC B105	Frameshift	d4		AAAG	10	B	1529	434
CRC B124	Frameshift	d4		AAAC	11	F	3034	936
CRC B135	Frameshift	i4		TTAG	10	B	1418	397
CRC B150	Frameshift	d2		TT	11	W	6503	2092
CRC B160	Frameshift	d5		TAACT	11	X	6630	2134
CRC B169	Frameshift	d1		G	11	A	2157	643
CRC B196	Frameshift	d2		TG	11	Y	6819	2197
CRC B211	Frameshift	d2		TT	11	W	6503	2092
CRC B245	Frameshift	d2		TC	11	X	6690	2154
CRC B247	Frameshift	i4		AGAT	23		9208	2994
IARC3594	Frameshift	d2		CT	11		5946	1906
IARC2932	Frameshift	d6		TTTCGg	13		7231	2334

TABLE 3

	Number of FBCs	Number of FBCs <50	Number of OvCs	Number of MBCs	LOD score at BRCA1	LOD score at BRCA2	BRCA2 mutation
IARC 2932	15	10	0	0	-2.38	3.01	CCC.TTT.CGgtaa
IARC 3594	6	5	0	0	nd	nd	CAT.AAC.TCT.CTA
CRC B211	5	3	4	0	-0.48	0.49	AGT.CTT.CAC
CRC B196	17	12	0	0	-2.21	0.92	AAA.ACT.GAA.ACT
Montreal 681	3	2	0	1	nd	nd	GCA.AGT.GGA
Montreal 440	2	2	0	2	nd	nc	GAT.AAA.CAA.GCA

References:

The references cited below and in the above description are all incorporated by reference in their entirety.

1. Hall et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science, 250(4988):1684-1689, 1990.

2. Malkin et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science, 250(4985):1233-1238, 1990.

3. Miki et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science, 266(5182):66-71, 1994.

4. Wooster et al. A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenstein syndrome. Nat. Genet., 2(2):132-134, 1992.

5. Wooster et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science, 265(5181):2088-2090, 1994.

6. Frohman et al. Rapid amplification of full length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. PNAS USA, 85:8998-9002, 1988.

7. Wooster et al. Identification Of the Breast-Cancer Susceptibility Gene BRCA2. Nature, 378:789-792, 1995.

8. Beaudet and Tsui. A Suggested Nomenclature for Designating Mutations. Hum. Mutat., 2(4):245-8, (1993).

9. Thorlacius et al. A Single BRCA2 Mutation In Male and Female Breast-Cancer Families From Iceland With Varied Cancer Phenotypes. Nature Genetics, 13:117-119, 1996.

10. Phelan et al. Mutation Analysis Of the BRCA2 Gene In 49 Site-Specific Breast-Cancer Families. Nature Genetics, 13:120-122, 1996.

11. Easton et al. Genetic Linkage Analysis in Familial Breast and Ovarian Cancer. Am. J. Hum. Genet., 52:718-722, 1993.

12. Gayther et al. Germline Mutations Of the Brca1 Gene In Breast and Ovarian-Cancer Families Provide Evidence For a Genotype-Phenotype Correlation. Nature Genetics, 11:428-433, 1995.

13. Holt et al. Growth-Retardation and Tumor-Inhibition by BRCA1. Nature Genetics, 12:298-302, 1996.

14. Gayther et al. Rapid Detection Of Regionally Clustered Germ-Line BRCA1 Mutations By Multiplex Heteroduplex Analysis. Am. J. Human Gen., 58:451-456, 1996.

15. Tavtigian et al. The Complete BRCA2 Gene and Mutations In Chromosome 13q-Linked Kindreds. Nature Genetics, 12:333-337, 1996.

16. Couch et al. BRCA2 Germline Mutations In Male Breast-Cancer Cases and Breast-Cancer Families. Nature Genetics, 13:123-125, 1996.

17. Neuhausen et al. Recurrent BRCA2 6174delT Mutations In Ashkenazi Jewish Women Affected By Breast-Cancer. Nature Genetics, 13:126-128, 1996.

18. Spirio et al. Alleles of the APC gene: an attenuated form of familial polyposis. Cell, 75:951-957, 1993.

19. Olschwang et al. Restriction of ocular fundus lesions to a specific subgroup of APC mutations in adenomatous polyposis coli patients. Cell, 75:959-968, 1993.

20. Bork et al. Internal Repeats In the BRCA2 Protein-Sequence. Nature Genetics, 13:22-23, 1996.

21. Cowley, S., Paterson, H.F., Kemp, P. and Marshall, C.J., Cell, 77:841-852, 1994.

22. Marais, R., Light, Y., Paterson, H.F. and Marshall, C.J., EMBO J., 14:3136-3145, 1995.

23. Hynes, G. et al, FASEB J. 10:137-147, 1996.

Claims:

1. A nucleic acid molecule comprising a part of the BRCA2 gene as set out in figures 1 or 2, or alleles thereof.

2. A nucleic acid molecule comprising the full length coding sequence or complete BRCA2 gene as obtainable by:

(a) using the nucleic acid sequences shown in figures 1 or 2 to construct probes for screening cDNA or genomic libraries, sequencing the positive clones obtained, and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained;

(b) using the sequences shown in figures 1 or 2 to obtain oligonucleotides for priming BRCA2 nucleic acid fragments, these oligonucleotides being used in conjunction with oligonucleotides designed to prime from a cloning vector, to amplify by PCR nucleic acid fragments in a library that contains fragments of the BRCA2 sequence, sequencing the amplified fragments to obtain the BRCA2 sequence between known parts of the sequence and the cloning vector, and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained; and/or,

(c) using rapid amplification of cDNA ends (RACE), by synthesizing cDNAs from a number of different RNAs, the cDNAs being ligated to an oligonucleotide linker, and amplifying by PCR the BRCA2 cDNAs using one primer that primes from the BRCA2 cDNA sequence of figure 1 and a second primer that primes from the oligonucleotide linker, sequencing the amplified nucleic acid and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained.

3. A nucleic acid molecule comprising a part of the BRCA2 gene as set out in figure 4 or alleles thereof.

4. A nucleic acid molecule comprising the full length coding sequence or complete BRCA2 gene as obtainable by:

(a) using the nucleic acid sequences shown in figures 1, 2 or 4 to construct probes for screening cDNA or genomic libraries, sequencing the positive clones obtained, and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained;

(b) using the sequences shown in figures 1, 2 or 4 to obtain oligonucleotides for priming BRCA2 nucleic acid fragments, these oligonucleotides being used in conjunction with oligonucleotides designed to prime from a cloning vector, to amplify by PCR nucleic acid fragments in a library that contains fragments of the BRCA2 sequence, sequencing the amplified fragments to obtain the BRCA2 sequence between known parts of the sequence and the cloning vector, and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained; and/or,

(c) using rapid amplification of cDNA ends (RACE), by synthesizing

cDNAs from a number of different RNAs, the cDNAs being ligated to an oligonucleotide linker, and amplifying by PCR the BRCA2 cDNAs using one primer that primes from the BRCA2 cDNA sequence of figures 1 or 4 and a second primer that primes from the oligonucleotide linker, sequencing the amplified nucleic acid and repeating this process to assemble the full length BRCA2 sequence from the sequences thus obtained.

5. A nucleic acid molecule which is an allele or variant of a BRCA2 nucleic acid molecule as obtainable in claim 2 or claim 4.

6. A nucleic acid molecule which has a nucleotide sequence encoding a BRCA2 polypeptide including the amino acid sequence set out in figure 3.

7. A nucleic acid molecule which has a nucleotide sequence encoding a BRCA2 polypeptide including the amino acid sequence set out in figure 5.

8. A nucleic acid molecule which has a nucleotide sequence encoding a BRCA2 polypeptide including the amino acid sequence set out in figure 7.

9. A nucleic acid molecule which has a nucleotide sequence encoding a polypeptide which is a variant, derivative or allele of a BRCA2 polypeptide including the amino acid sequence set out in any one of figure 3, 5 or 7.

10. The nucleic acid of claim 9 encoding a polypeptide having at least 80% sequence homology to the BRCA2 polypeptide including the amino acid sequence set out in any one of figure 3, 5 or 7.

11. The nucleic acid molecule of claim 9 or claim 10 wherein the molecule has one or more of the mutations set out in table 1.

12. The nucleic acid molecule of claim 11 having a 6174delT mutation.

13. The nucleic acid molecule of claim 11 having a 6503delTT mutation.

14. A nucleic acid molecule which has a nucleotide sequence encoding a fragment or active portion of a BRCA2 polypeptide including the amino acid sequence set out in any one of figure 3, 5, or 7.

15. The nucleic acid molecule of any one of claims 1 to 14 further comprising all or a part the BRCA2 promoter region, the nucleic acid sequence of which is set out in figure 6.

16. A replicable vector comprising nucleic acid of any one of claims 1 to 14 operably linked to control sequences to direct its expression.

17. The vector of claim 16 further comprising the nucleic acid of claim 15 operably linked to promote the expression of the nucleic acid encoding the BRCA2 polypeptide.

18. Host cells transformed with the vector of claim 16 or claim 17.

19. A method of producing a BRCA2 polypeptide comprising culturing the host cells of claim 18 so that BRCA2 polypeptide is produced.

20. The method of claim 19 comprising the further step of recovering the polypeptide produced.

21. The nucleic acid molecule of any one of claims 1 to 15, or its complement, further comprising a label.

22. A nucleic acid molecule of any one of claims 1 to 15 for use in a method of medical treatment.

23. A substance which is a BRCA2 polypeptide encoded by the nucleic acid of claim 2 or claim 4.

24. A substance which is a BRCA2 polypeptide including the amino acid sequence set out in figure 3.

25. A substance which is a BRCA2 polypeptide including the amino acid sequence set out in figure 5.

26. A substance which is a BRCA2 polypeptide including the amino acid sequence set out in figure 7.

27. A substance which is a polypeptide having at least 80% sequence homology to the BRCA2 polypeptide including the amino acid sequence set out in any one of figure 3, 5 or 7.

28. A substance which is a polypeptide which is a variant, derivative or allele of a BRCA2 polypeptide of any one of claims 22 to 27.

29. A substance which is a fragment or active portion or functional mimetic of a BRCA2 polypeptide including the amino acid sequence of any one of figures 3, 5, or 7.

30. The substance of any one of claims 22 to 29 further comprising a label.

31. The substance of any one of claims 22 to 30 for use in a method of

medical treatment.

32. An antibody capable of specifically binding to a BRCA2 polypeptide of any one of claims 22 to 30.

33. The antibody of claim 32 further comprising a label.

34. A pharmaceutical composition comprising a substance of any one of claims 22 to 30.

35. A pharmaceutical composition comprising an antibody of claim 32 or claim 33.

36. The pharmaceutical composition of claim 34 or claim 35 additionally comprising a pharmaceutically acceptable carrier.

37. A method of diagnosing a susceptibility or predisposition to cancer in a patient by analysing a sample from the patient for the BRCA2 gene or the polypeptide encoded by it, the method comprising:

(a) comparing the sequence of nucleic acid in the sample with the BRCA2 nucleic acid sequence as set out in figure 1, 2, 4 or 7 to determine whether the sample from the patient contains mutations; or,

(b) determining the presence in the sample of the polypeptide encoded by the BRCA2 gene as set out in the partial sequences of figures 3 and 5 or the full length sequence of figure 7 and, if present, determining whether the polypeptide is full length, and/or is mutated, and/or is expressed at the normal level; or,

(c) using DNA fingerprinting to compare the restriction pattern produced when a restriction enzyme cuts a sample of nucleic acid from the patient with the restriction pattern obtained from normal BRCA2 gene comprising the sequence set out in figures 1, 2, 4 or 7 or from known mutations thereof; or,

(d) using a specific binding member capable of binding to a BRCA2 nucleic acid sequence as set out in figures 1, 2, 4 or 7 (either a normal sequence or a known mutated sequence), the specific binding member comprising nucleic acid hybridisable with the BRCA2 sequence, or substances comprising an antibody domain with specificity for a native or mutated BRCA2 nucleic acid sequence or the polypeptide encoded by it, and detecting the binding of the specific binding member to its binding partner by means of a label; or,

(e) using PCR involving one or more primers based on normal BRCA2 gene sequence set out in figures 1, 2, 4 or 7, or mutated forms thereof, to screen for normal or mutant BRCA2 gene in a sample from a patient.

38. The method of claim 37 wherein the patient sample is analysed by

single stranded conformation polymorphism (SSCP) and/or by a protein truncation (PTT) assay.

39. A method of identifying a target nucleic acid molecule in a test sample using a nucleic acid probe having a portion of the sequence shown in any one of figures 1, 2, 4 or 7 or a complementary sequence thereof, the method comprising contacting the probe and the test sample under hybridising conditions and observing whether hybridisation takes place.

40. The method of claim 39 wherein the probe is used to identify a BRCA2 nucleic acid sequence or a mutant allele thereof.

41. The method of claim 39 wherein the probe is used to identify BRCA2 nucleic acid of other species.

42. A method of determining the presence of one or more mutations in a sample of nucleic acid from a patient using one or more allele specific nucleic acid probes having the sequence or a portion of the sequence set out of figures 1, 2, 4 or 7 or a complementary sequence thereof, the method comprising contacting the probe and the test sample under hybridising conditions and observing whether hybridisation takes place.

43. The method of claim 42 wherein a plurality of allele specific probes are used, the probes being immobilised in an array on a solid support, the hybridisation of the nucleic acid sample to the probes being detected by means of a label.

44. The method of claim 42 or claim 43 wherein the mutations include one or more of those set out in table 1.

45. The use of a nucleic acid molecule of any one of claims 1 to 15 in the preparation of a medicament for treating cancer.

46. The use of claim 45 wherein the cancer is female breast cancer, male breast cancer, ovarian cancer, prostate cancer, colorectal cancer, ocular melanoma or leukaemia.

47. The use of claim 45 or claim 46 wherein the nucleic acid molecule is an antisense oligonucleotide capable of hybridising to the complementary sequence of BRCA2 nucleic acid, pre-mRNA or mature mRNA so that the expression of the BRCA2 nucleic acid is reduced or prevented.

48. The use of claim 45 or claim 46 wherein the use of the nucleic acid is in a method of gene therapy.

49. The use of a nucleic acid sequence of any one of claims 1 to 15 in the design of primers for use in the polymerase chain reaction.

50. The use of a nucleic acid sequence of any one of claims 1 to 15 in the design of a nucleic acid probe for detecting the presence of mutations in a nucleic acid sample from a patient.

51. The use of nucleic acid encoding all or a functional part of the BRCA2 promoter region set out in figure 6 in screening for substances which modulate the expression of nucleic acid under the control of the promoter.

52. A polypeptide having a BRCA2 amino acid motif set out in table 4, the motif being conserved between the human BRCA2 sequence set out in any one of figures 3, 5 or 7 and the sequence of another species.

53. Use of a substance of any one of claims 23 to 30 in the preparation of a medicament for treating cancer.

54. The use of claim 53 wherein the cancer is female breast cancer, male breast cancer, ovarian cancer, prostate cancer, colorectal cancer, ocular melanoma or leukaemia.

55. The use of an antibody of claim 32 or claim 33 for determining the presence, amount or location in a cell of BRCA2 polypeptide or mutant forms thereof.

56. Use of substance of any one of claims 23 to 30 to screen for binding partners to the substance.

57. A method of screening for substances which mimic the activity of BRCA2 polypeptide or a portion thereof, the method comprising contacting the test substances with a BRCA2 specific binding partner and, and determining whether the test substances bind to the specific binding partner.

58. The method of claim 57 further comprising the testing the substances binding to the specific binding partner for activity as cancer therapeutics.

59. A method of screening for substances which affect or modulate the activity of the BRCA2 polypeptide of any one of claims 23 to 30, the method comprising contacting one or more test substances with the BRCA2 polypeptide in a reaction medium, testing the activity of the treated BRCA2 polypeptide and comparing that activity with the activity of the BRCA2 polypeptide in comparable reaction medium untreated with the test substance

or substances.

60. A kit for detecting mutations in the BRCA2 gene associated with a susceptibility to cancer, the kit comprising one or more nucleic acid probes capable of specifically binding a mutated BRCA2 nucleic acid sequence.

61. A kit for detecting mutations in the BRCA2 gene associated with a susceptibility to cancer, the kit comprising one or more antibodies capable of specifically binding a mutated BRCA2 nucleic acid sequence.

62. A kit comprising at least one oligonucleotide primer having a sequence corresponding to or complementary to a portion of the nucleic acid sequence set out in any one of figures 1, 2, 4 or 7 for use in amplifying a BRCA2 nucleic acid sequence or an allele thereof.

63. The kit of claim 62 wherein the primers are for detecting mutations in the BRCA2 gene by single stranded conformation polymorphism (SSCP) and/or by a protein truncation (PTT) test.

64. A kit for determining the presence of one or more mutations in a sample of nucleic acid from an individual, the kit comprising:

(a) a solid support having immobilised thereon one or more allele specific nucleic acid probes having sequences corresponding to portions of the sequence set out of figures 1, 2, 4 or 7 or a complementary sequence thereof and/or one or more antibodies capable of specifically binding a mutated BRCA2 nucleic acid sequence; and,

(b) a label for marking the presence of sample nucleic acid hybridised to the probe(s) or antibodies, or to probes or antibodies not hybridised to sample nucleic acid.

65. The kit of claim 64 wherein the label is adapted for binding to the nucleic acid sample prior to contacting the sample with the support.

66. The kit of claim 64 wherein the label is associated with a developing agent, the developing agent being capable of binding to sample nucleic acid hybridised to the probe(s) or antibodies, or to probes or antibodies not hybridised to the sample nucleic acid.

67. A chimeric animal having an normal BRCA2 allele.

68. A chimeric animal having a BRCA2 allele having one or more of the mutations set out table 1.

The Patent Office

Application No:	GB 9624453.8	**Examiner:**	Colin Sherrington
Claims searched:	1 to 68	**Date of search:**	28 February 1997

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.O): C3H(HB7P,HB7M)

Int Cl (Ed.6): C07K 14/82; C12Q 1/68; G01N 33/574

Other: ONLINE: WPI,CLAIMS,DIALOG/BIOTECH

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	Science 1994,265,2088-2090 -Richard Wooster *et al.* "Localization of a Breast Cancer Susceptibility Gene, *BRCA2*, to Chromosome 13q12-13"	1, 3
P,A	Nature Genetics 1996,12,333-337 -S.V.Tavtigian *et al.* "The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds"	1, 3
P,A	Oncogene 1996,13,213-221 -ANK Jacob *et al.* "Isolation of expressed sequences that include a gene for familial breast cancer (BRCA2) and other novel transcripts from a five megabase region on chromosome 13q12"	1 to 15, 23 to 29, 32
P,A	Proc.Natl.Acad.Sci.USA 1996,93,690-694 -Stuart G.Fischer *et al.* "A high-resolution annotated physical map of the human chromosome 13q12-13 region containing the breast cancer susceptibility locus *BRCA2*"	1, 3

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.